

**Verzerrte Glaubwürdigkeitsbeurteilung im Schulkontext:
Effekt unbewusster Urteilsziele und des
Verarbeitungsmodus auf die Glaubwürdigkeitsbeurteilung
von Schülerinnen und Schülern**

Inauguraldissertation

zur

Erlangung des akademischen Grades einer Doktorin
der Sozialwissenschaften der Universität Mannheim

vorgelegt von

Tamara Marksteiner

Dekan:

Prof. Dr. Michael Diehl

Vorsitzende der Prüfungskommission: **Prof. Dr. Dagmar Stahlberg**

Erstgutachter: **Prof. Dr. Oliver Dickhäuser**

Zweitgutachter: **PD Dr. Marc-André Reinhard**

Tag der Disputation: **23.05.2013**

Erklärung

Hiermit erkläre ich, dass ich die vorgelegte Dissertation selbstständig und nur mit Hilfen angefertigt habe, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht.

Unterschrift

Datum

Inhaltsverzeichnis

Zusammenfassung	1
1 Einleitung	1
1.1 Ziel der Arbeit	1
1.2 Glaubwürdigkeit, Täuschung und Lüge	2
1.2.1 Definition von Glaubwürdigkeit, Täuschung und Lüge.....	3
1.2.2 Besonderheiten von Glaubwürdigkeitsurteilen	7
1.3 Täuschung im Schulkontext	10
1.3.1 Subjektive Häufigkeit von Täuschungssituationen im Schulkontext	10
1.3.2 Häufigkeit von Täuschungssituationen im Schulkontext.....	11
1.4 Erkennen von Lüge und Wahrheit	12
1.4.1 Durchschnittliche Trefferquote.....	12
1.4.2 Trefferquote von Laien und Experten.....	13
1.4.3 Trefferquote und Urteilsverzerrungen.....	15
1.5 Soziale Informationsverarbeitung.....	17
1.5.1 Modi der Informationsverarbeitung	18
1.5.2 Informationsverarbeitung und Glaubwürdigkeitsurteile	24
1.6 Verzerrungen bei der Informationsverarbeitung.....	27
1.6.1 Motivation, Ziele und verzerrte Informationsverarbeitung	28
1.6.2 Verzerrte Informationsverarbeitung und Glaubwürdigkeitsurteile	35
1.7 Modell der verzerrten Informationsverarbeitung.....	45
2 Pilotstudie 1: Der Einfluss von Urteilszielen mit angenommenen Urteilskonsequenzen auf Glaubwürdigkeitsurteile	50
2.1 Fragestellung und Hypothesen	51
2.2 Methode	52
2.2.1 Versuchsablauf	52

2.2.2	Versuchsmaterial und Vortest des Gesprächs	53
2.2.3	Versuchsdesign	57
2.2.4	Stichprobe	57
2.3	Ergebnisse	58
2.3.1	Manipulationscheck	58
2.3.2	Glaubwürdigkeitsurteil	59
2.4	Diskussion	60
3	Pilotstudie 2: Der Einfluss von unbewussten Urteilszielen mit realen Urteilskonsequenzen auf Glaubwürdigkeitsurteile	61
3.1	Fragestellung und Hypothesen	61
3.2	Methode	62
3.2.1	Versuchsablauf	62
3.2.2	Versuchsmaterial	64
3.2.3	Versuchsdesign	64
3.2.4	Stichprobe	64
3.3	Ergebnisse	65
3.3.1	Glaubwürdigkeitsurteil	65
3.3.2	Binäres Glaubwürdigkeitsurteil	65
3.3.3	Argumente allgemein	66
3.3.4	Argumente für Urteil „wahr“ vs. „erfunden“	67
3.4	Diskussion	69
4	Experiment 1: Einfluss von unbewussten Urteilszielen und Verarbeitungsmodus auf Glaubwürdigkeitsurteile	71
4.1	Fragestellung und Hypothese	71
4.2	Methode	72
4.2.1	Versuchsablauf	72
4.2.2	Versuchsmaterial	74
4.2.3	Versuchsdesign	75

4.2.4	Stichprobe	75
4.3	Ergebnisse	75
4.3.1	Glaubwürdigkeitsurteil	75
4.3.2	Binäres Glaubwürdigkeitsurteil	76
4.3.3	Argumente allgemein	77
4.3.4	Argumente für Urteil „wahr“ vs. „erfunden“	78
4.4	Diskussion	80
5	Experiment 2: Einfluss von unbewussten Urteilszielen und Verarbeitungsmodus auf die Akkuratheit von Glaubwürdigkeitsurteilen	83
5.1	Fragestellung und Hypothesen	83
5.2	Methode	86
5.2.1	Versuchsablauf	86
5.2.2	Versuchsmaterial	89
5.2.3	Versuchsdesign	93
5.2.4	Stichprobe	93
5.2.5	Abhängige Variablen	94
5.3	Ergebnisse	95
5.3.1	Wahrheitsverzerrung	95
5.3.2	Trefferquote	96
5.3.3	Trefferquote bei wahren und erfundenen Aussagen	98
5.4	Diskussion Experiment 2	99
6	Mehrebenenanalysen und Glaubwürdigkeitsforschung	101
6.1	Grundlagen der Mehrebenenanalyse	102
6.1.1	Genestete und nicht-genestete Modelle	102
6.1.2	Lineare, logistische und Mehrebenen-Regression	104
6.2	Mehrebenenanalysen in der Glaubwürdigkeitsforschung	108

7	Experiment 3: Einfluss von unbewussten Urteilszielen und Verarbeitungsmodus auf die Akkuratheit von Glaubwürdigkeitsurteilen unter der Berücksichtigung von Schülermerkmalen.....	111
7.1	Fragestellung und Hypothesen	111
7.2	Methode	114
7.2.1	Versuchsablauf	114
7.2.2	Versuchsmaterial	115
7.2.3	Versuchsdesign	120
7.2.4	Stichprobe	120
7.2.5	Abhängige Variablen und Vorgehen bei der Ergebnisauswertung.....	121
7.3	Ergebnisse	122
7.3.1	Hypothesentestung anhand logistischer Mehrebenenanalyse	122
7.3.2	Hypothesentestung anhand von ANOVAs	128
7.4	Diskussion Experiment 3.....	132
8	Abschlussdiskussion	139
8.1	Effekt von unbewussten Urteilszielen und Verarbeitungsmodus auf Glaubwürdigkeitsbeurteilung von Schülern	139
8.1.1	Einfluss auf die Glaubwürdigkeitsbeurteilung	139
8.1.2	Einfluss auf die Akkuratheit von Glaubwürdigkeitsbeurteilungen.....	141
8.1.3	Auswertung der Daten mittels Mehrebenenregressionen	142
8.2	Praktische Implikationen	144
8.3	Ausblick.....	146
9	Literatur	149
	Anhang.....	161

Zusammenfassung

In der vorliegenden Arbeit wurde bei (angehenden) Lehrkräften der Einfluss von Urteilszielen und des Modus, in dem urteilsrelevante Informationen verarbeitet werden, auf den Prozess der Glaubwürdigkeitsbeurteilung von Schülern¹, auf das Glaubwürdigkeitsurteil und auf die Akkuratheit dieser Urteile untersucht.

Bei der Hypothesengenerierung wurde insbesondere auf Annahmen eines Zwei-Prozess-Modells der sozialen Informationsverarbeitung, des Heuristic-Systematic-Models (Chen & Chaiken, 1999), zurückgegriffen. Es wurde angenommen, dass (angehende) Lehrkräfte aufgrund von Urteilszielen Informationen zur Bildung eines Glaubwürdigkeitsurteils über einen Schüler in verzerrter Art und Weise verarbeiten und dies zu einer Verzerrung des Glaubwürdigkeitsurteils führt. Weiterhin wurde angenommen, dass bei hoher Verarbeitungsmotivation Informationen verzerrt systematisch verarbeitet werden, bei niedriger Verarbeitungsmotivation dagegen verzerrt heuristisch. Unabhängig vom Urteilsziel soll bei systematischer Verarbeitung eine höhere Akkuratheit der Urteile resultieren als bei heuristischer Verarbeitung.

Durchgeführt wurden zwei Pilotstudien und drei Experimente. In den Pilotstudien 1 und 2 wurde die Manipulation des Urteilsziels vorgetestet. Dabei zeigte sich, dass unbewusste Urteilsziele, welche *mit* tatsächlichen Konsequenzen für den Urteilenden verbunden sind, einen Einfluss auf die

¹ In der vorliegenden Arbeit wird nur die maskuline Bezeichnung verwendet, um den Lesefluss zu erleichtern. Gemeint sind sowohl männliche auch als weibliche Personen, sofern nicht ein Geschlecht explizit hervorgehoben wird.

Urteilsverzerrung haben, nicht jedoch Urteilsziele *ohne* tatsächliche Konsequenzen.

In Experiment 1 wurde untersucht, wie sich Urteilsziel und Verarbeitungsmodus auf den Prozess der Glaubwürdigkeitsbeurteilung von Schülern auswirken. Die Ergebnisse deuten darauf hin, dass ein Urteilsziel unabhängig vom Verarbeitungsmodus in einer Urteilsverzerrung endet und dass das Urteilsziel in Abhängigkeit des Verarbeitungsmodus zu einer verzerrten heuristischen oder systematischen Verarbeitung führt. Die Ergebnisse von Experiment 2 zeigen, dass Lehrkräfte unabhängig vom Verarbeitungsmodus einer Urteilsverzerrung unterliegen und dass unabhängig vom Urteilsziel eine höhere Akkuratheit der Glaubwürdigkeitsurteile bei systematischer (vs. heuristischer) Verarbeitung erzielt wird. Experiment 3 repliziert die Befunde von Experiment 2 und zeigt darüber hinaus, dass Variablen höherer Ebene, d.h. Schülermerkmale (z.B. die Länge der Schülerantwort), die Urteilsverzerrung beeinflussen. Um Variablen höherer Ebene einzubeziehen, wurde erstmals in der Glaubwürdigkeitsforschung ein nicht-hierarchisches, logistisches Mehrebenen-Regressionsmodell zur Modellierung und Analyse der Daten verwendet.

1 Einleitung

1.1 Ziel der Arbeit

Damit Lehrkräfte ihren Lehr- und Erziehungsauftrag (vgl. Terhart, 2005, Standards für die Lehrerbildung) erfolgreich erfüllen können, ist es von großer Bedeutung, dass sie ihre Schüler akkurat beurteilen. Eine wichtige Komponente ist dabei die Glaubwürdigkeitsbeurteilung, z.B. wenn der Verdacht aufkommt, dass ein Schüler bei einer Klassenarbeit abgeschrieben hat. Bisherige Studien (Reinhard, Dickhäuser, Marksteiner, & Sporer, 2011; Vrij, Akehurst, Brown, & Mann, 2006) zeigten, dass Lehrkräfte bei der Beurteilung der Glaubwürdigkeit von Schülern dazu tendieren, deren Aussagen als „wahr“ einzustufen (d.h. ihr Urteil zu *verzerren*), weil sie mit diesem Urteil ein bestimmtes *Ziel* verfolgen, z.B. eine falsche Beschuldigung zu vermeiden (Hurst-Wagner & Oswald, 2011; Marksteiner, Reinhard, Lettau, & Dickhäuser, 2013). Zudem wird die Beurteilung der Glaubwürdigkeit von der *Motivation* des Urteilers (Informationen zu verarbeiten) beeinflusst, welche ihrerseits den *Verarbeitungsmodus* bestimmt (Chaiken, 1980).

Die vorliegende Arbeit geht daher der Frage nach, auf welche Art und Weise (angehende) Lehrkräfte Glaubwürdigkeitsurteile über Schüler in Abhängigkeit von Urteilsziel und Verarbeitungsmodus bilden und wie sich diese Faktoren letztlich auf die Verzerrung und die Akkuratheit von Glaubwürdigkeitsurteilen auswirken. Die theoretische Grundlage dieser Untersuchung stellt ein allgemeines Modell der sozialen Informationsverarbeitung, das *Heuristic-Systematic Model* (HSM; Chaiken, 1980), dar. Während bisherige Untersuchungen einen varianzanalytischen

Ansatz wählen (vgl. C. F. Bond & DePaulo, 2006), besteht die Besonderheit der vorliegenden Arbeit darin, dass ein *nicht-hierarchisches, logistisches Mehrebenen-Regressionsmodell* (vgl. Hox, 2010; Twisk, 2011) zur Modellierung und Analyse der Daten verwendet wird.

In einem ersten Schritt werden dazu die Begriffe „Glaubwürdigkeit“, „Täuschung“ sowie „Lüge“ definiert und die Besonderheiten von Glaubwürdigkeitsurteilen im Schulkontext dargestellt (dazu Kapitel 1.2). Sodann werden sowohl Situationen, in denen Lehrkräfte die Beurteilung der Glaubwürdigkeit ihrer Schüler für relevant halten, als auch deren Häufigkeit im Schulalltag erörtert (dazu Kapitel 1.3), sowie bisherige Befunde zum Erkennen von Lüge und Wahrheit dargestellt (dazu Kapitel 1.4). In einem weiteren Schritt wird ausführlich auf das HSM (dazu Kapitel 1.5) und darauf aufbauend auf Faktoren eingegangen, welche die Informationsverarbeitung verzerren (dazu Kapitel 1.6). Abschließend wird ein Prozessmodell der verzerrten Informationsverarbeitung bei Glaubwürdigkeitsbeurteilungen vorgestellt (dazu Kapitel 1.7).

1.2 Glaubwürdigkeit, Täuschung und Lüge

Im folgenden Abschnitt wird zunächst der Begriff „Glaubwürdigkeit“ definiert und anschließend den Begriffen „Täuschung“ und „Lüge“ gegenübergestellt. Dabei wird auf die englischen Übersetzungen der Begriffe eingegangen, da diese von der Forschung und in den zum größten Teil englischsprachigen Fachartikeln verwendet werden. Des Weiteren werden die Besonderheiten von Glaubwürdigkeitsurteilen und insbesondere im Schulkontext betrachtet.

1.2.1 Definition von Glaubwürdigkeit, Täuschung und Lüge

Glaubwürdigkeit. Einen sehr spezifischen Überblick über die Möglichkeiten der Definition des Begriffs Glaubwürdigkeit gibt Köhnken in seinem Buch über Glaubwürdigkeit (1990). Ins Englische wird der Begriff Glaubwürdigkeit laut Köhnken meistens mit „credibility“ übersetzt. Bei der Definition des Begriffs Glaubwürdigkeit ist es hilfreich, zwischen dem *Kommunikator*, d.h. der Person, die eine Information übermittelt, und dem *Rezipienten*, d.h. der Person, die eine Information erhält, zu unterscheiden und zwischen tatsächlich *wahren* und tatsächlich *falschen* Informationen zu differenzieren. So definiert Köhnken *Glaubwürdigkeit* als

„...Wahrheitsvorsatz‘ bzw. Abwesenheit von absichtlichen Täuschungen des Kommunikators“ (1990, S. 3).

Er erläutert weiter, dass Glaubwürdigkeit dann vorliege,

„...wenn ein Kommunikator an einen Rezipienten eine Information vermittelt, von der der Kommunikator glaubt, dass sie zutreffend ist, wenn also der Kommunikator keine Täuschungsabsicht hat“ (S. 4).

Diese Erläuterungen beziehen sich folglich mehr darauf, unter welchen Umständen ein Kommunikator tatsächlich glaubwürdig ist –nämlich dann, wenn er selbst glaubt, dass die von ihm übermittelte Information zutreffend ist, unabhängig davon, ob die Information tatsächlich richtig oder falsch ist. Betont wird hier laut Köhnken damit die *Intentionalität* des Kommunikators. Somit

kann der Kommunikator auch irrtümlich und ohne Vorsatz eine falsche Information vermitteln, ohne unglaubwürdig zu sein.

Nähert man sich dem Begriff der Glaubwürdigkeit hingegen aus Sicht des Rezipienten, steht die durch diesen *wahrgenommene* Glaubwürdigkeit des Kommunikators im Zentrum. So kann ein Kommunikator eine *richtige* Information vermitteln (also tatsächlich glaubwürdig sein), aber vom Rezipienten dennoch als *unglaubwürdig* wahrgenommen werden. Umgekehrt kann ein Kommunikator eine *falsche* Information vermitteln (also tatsächlich unglaubwürdig sein), aber vom Rezipienten als *glaubwürdig* wahrgenommen werden.

Glaubwürdigkeit und Täuschung. Köhnken (1990) geht davon aus, dass Glaubwürdigkeit ebenso wie Täuschung ein Kommunikationsphänomen ist, d.h. dass an einer Täuschung immer zwei Personen (Kommunikator und Rezipient) beteiligt sind. Köhnken definiert „Täuschung“ in der Folge als

„...Verhaltensweisen, mit denen bei anderen Personen ein Eindruck erzeugt werden soll, von dem der Handelnde annimmt, dass er falsch ist“ (1990, S. 3).

Bei einer Täuschung vermittelt also der Kommunikator absichtlich eine falsche Information an den Rezipienten. Laut Köhnken (1990) ist eine Täuschung mittels zahlreicher Verhaltensweisen möglich. So kann *verbal* getäuscht werden, indem absichtlich eine falsche Aussage gemacht wird (verbale Täuschung), aber auch *nonverbal*, indem z.B. trotz besseren Wissens durch ein Kopfnicken eine Frage bejaht wird (nonverbale Täuschung).

Diese Definition von „Täuschung“ deckt sich weitgehend mit der Definition von Glaubwürdigkeit (der Vermittlung von Information, die der Kommunikator für zutreffend hält). Wie auch beim Begriff der Glaubwürdigkeit ist bei einer Täuschung die Intentionalität von entscheidender Bedeutung: Der Kommunikator vermittelt vorsätzlich nicht zutreffende Informationen an den Rezipienten.

Der Begriff der Täuschung kann ins Englische entweder mit „deception“ oder „cheating“ übersetzt werden, wobei der Begriff des cheating eher im akademischen Kontext verwendet wird. Da für die vorliegende Arbeit Täuschungen im akademischen Kontext von Bedeutung ist, wird hier auf den Begriff „cheating“ näher eingegangen. „Cheating“ kann definiert werden als

„...unmoral and socially undesired behavior that can lead to lying about the behavior in question in order to avoid punishment“

(Marksteiner & Reinhard, 2013),

d.h. als unmoralisches und sozial unerwünschtes Verhalten, das – um Strafe zu entgehen – zum Lügen führen kann. Der Begriff kann ins Deutsche u.a. als „Mogeln“ zurückübersetzt werden, wobei laut Pegels (1997) „cheating“ ebenso dem Betrügen, Lügen, Schummeln und Pfuschen entsprechen kann. Für einen einzigen Begriff – „cheating“ – gibt es somit eine Vielzahl an Übersetzungsmöglichkeiten ins Deutsche, wobei durch die jeweilige deutsche Übersetzung der Begriff auch eine andere Bedeutung erhält (Pegels, 1997). Für die vorliegende Arbeit ist u.a. „cheating“ im Sinne von Mogeln von Bedeutung, wobei Mogeln wiederum definiert werden kann als

„...kleiner Betrug‘, bei dem es darum geht, sich selbst oder anderen heimlich, mit unerlaubten Mitteln einen Vorteil zu verschaffen“
(Pegels, 1997, S. 51).

Glaubwürdigkeit, Täuschung und Lüge. Inhaltlich grenzt Köhnken das Konstrukt der Lüge von „Täuschung“ und „Glaubwürdigkeit“ ab, indem er die Lüge als die verbale Komponente der Täuschung (1990, S. 3) betrachtet. Auch Bond und das Global Deception Research Team (C. F. Bond, 2006) definieren die Lüge als

„...intentional false statement“ (absichtliche Falschaussage; S. 62).

Ein grundlegender Unterschied zwischen den Konstrukten „Glaubwürdigkeit“ und „Lüge“ ist, dass in der bisherigen Forschung das Konstrukt der Glaubwürdigkeit meistens metrisch, d.h. auf einer mehrstufigen Skala von 1 (unglaubwürdig) bis 10 (glaubwürdig), erfasst wurde (z.B. Marksteiner, Reinhard, Dickhäuser, & Sporer, 2012; Reinhard & Sporer, 2010). Ob ein Proband lügt oder die Wahrheit sagt, wird hingegen gewöhnlich anhand eines binären Urteils (wahr vs. erlogen) erfasst (vgl. C. F. Bond & DePaulo, 2006). Bei der Frage danach, ob eine Person lügt oder die Wahrheit sagt, wird damit strenggenommen das Konstrukt „Lüge“ erfasst.

Bei der Frage nach der Glaubwürdigkeit einer Person beeinflussen ebenfalls verbale Informationen bzw. der Nachrichteninhalt (wie z.B. Details) als auch nonverbale Informationen bzw. Quellenmerkmale (wie z.B. Handbewegung des Senders) das Urteil (vgl. Reinhard & Sporer, 2008, 2010). Laut der Metaanalyse von Hartwig und Bond (2011) berücksichtigen Rezipienten bei der Beantwortung der Frage sowohl verbale Informationen (z.B.

die logische Konsistenz der Aussage des Senders) als auch nonverbale Informationen (z.B. Bewegungen des Körpers des Senders). Somit werden sowohl bei der Frage, ob eine Person glaubwürdig ist, als auch bei der Frage, ob eine Person lügt oder die Wahrheit sagt, neben verbalen auch nonverbale Indikatoren zur Bildung des Urteils genutzt.

Zusammenfassung. Insgesamt lassen sich die Begriffe „Glaubwürdigkeit“, „Täuschung“ und „Lüge“ nur schwer voneinander abgrenzen (vgl. Köhnken, 1990). Diese Schwierigkeit betreffen sowohl die inhaltliche Abgrenzung der Konstrukte als auch die Abgrenzung im Rahmen der Erfassung der Konstrukte. Während *Glaubwürdigkeit* das Vermitteln einer Information ist, die der Kommunikator für zutreffend hält, kann *Täuschung* als das Vermitteln einer Information verstanden werden, die der Kommunikator für nicht zutreffend hält. Die Lüge wird dabei vom Begriff der Täuschung umfasst und ist die verbale Vermittlung von nicht-zutreffenden Informationen. Sowohl bei der Beurteilung der Glaubwürdigkeit eines Kommunikators als auch bei der Frage, ob ein Kommunikator lügt oder die Wahrheit sagt, verwendet der Rezipient nonverbale und verbale Informationen – wobei sich eine Lüge per definitionem lediglich auf verbal vermittelte Informationen bezieht.

1.2.2 Besonderheiten von Glaubwürdigkeitsurteilen

Besonderheiten von Glaubwürdigkeitsurteilen. Die Beurteilung der Glaubwürdigkeit kann zur Kategorie „Personenbeurteilung“ gezählt werden. Im Vergleich zu Urteilen über Gegenstände, z.B. die Einstellung gegenüber einem Anrufbeantworter (z.B. Chaiken & Maheswaran, 1994), und im Vergleich zu anderen Urteilen über Personen, z.B. die Einschätzung der akademischen

Fähigkeiten einer Person (z.B. Darley & Gross, 1983), beinhaltet die Beurteilung der Glaubwürdigkeit einer Person zudem einen *moralischen Aspekt* (vgl. Pizarro, 2010). Bei der Beurteilung der Glaubwürdigkeit einer Person wird zum einen die Frage in den Raum gestellt, ob deren Aussagen als glaubwürdig und vertrauenswürdig einzuschätzen sind. Zum anderen stellt sich der Urteilende die Frage, ob sich die zu beurteilende Person moralisch richtig oder falsch verhalten hat. Was als moralisches oder unmoralisches Verhalten angesehen wird, ist in manchen Fällen kodifiziert und in Gesetzen sowie Vorschriften schriftlich festgehalten.

In vielen Berufssparten müssen häufig Urteile darüber gefällt werden, ob eine andere Person die Wahrheit sagt oder lügt. So müssen z.B. Richter und Polizeibeamte die Aussagen von Angeklagten und Zeugen einstufen, Lehrer die Berichte ihrer Schüler und Sozialarbeiter die Angaben ihrer Klienten. In diesen Fällen wird den zu beurteilenden Personen häufig eine moralisch verwerfliche Handlung unterstellt (z.B. dem Angeklagten die Beteiligung an einem Verbrechen, dem Zeugen den Schutz des Täters und somit die Behinderung der Ermittlungsarbeiten, dem Schüler das Abschreiben bei einem Test und dem Klienten das Fälschen seines Lebenslaufs).

Besonderheiten von Glaubwürdigkeitsurteilen im Schulkontext. Zum moralischen Aspekt bei Glaubwürdigkeitsurteilen tritt im Schulkontext zum einen der *pädagogische Erziehungsaspekt* (vgl. Terhart, 2005, Standards für die Lehrerbildung) bei deviantem Verhalten und zum anderen die *Lehrer-Schülerbeziehung* hinzu. In Bezug auf den *pädagogischen Erziehungsaspekt* wird vom Lehrer erwartet, dass Täter zur Rechenschaft gezogen und Opfer

geschützt und entschädigt werden. In Bezug auf die *Lehrer-Schülerbeziehung* wird dagegen angenommen, dass sich die falsche Beschuldigung eines Schülers durch den Lehrer negativ auf das Vertrauensverhältnis und den daraus resultierenden Umgang zwischen Lehrer und Schüler auswirken kann (vgl. Marksteiner, Reinhard, et al., 2013). Ein Schüler, der z.B. bei einer sehr guten Testleistung vom Lehrer gefragt wird, ob er einen Spickzettel verwendet hat, obwohl dies nicht der Fall war, könnte die Frage des Lehrers als Anschuldigung derart betrachten, dass der Schüler eine sehr gute Leistung ohne Spickzettel nicht hätte erzielen können. Dies könnte sich wiederum negativ auf das schulische Selbstkonzept des Schülers und das zukünftige Verhalten bei Tests auswirken. Lehrer sollten somit die Konsequenzen von Falschbeschuldigungen in Betracht ziehen, bevor sie einen Schüler befragen. Die Annahme, dass Lehrkräfte sich der negativen Konsequenzen falscher Anschuldigungen tatsächlich bewusst sind, wird durch eine Untersuchung von Marksteiner, Reinhard, et al. (2013) unterstützt. Die Befunde dieser Untersuchungen deuten darauf hin, dass Lehrkräfte in Situationen, wie aggressivem Verhalten gegenüber Mitschülern, Diebstahl, Bullying und Vandalismus, aber auch bei dem Benutzen unerlaubter Hilfsmittel sowie unerlaubter Abwesenheit, es im Allgemeinen für wichtiger halten, keine Falschbeschuldigungen vorzunehmen, als das Fehlverhalten von Schülern aufzudecken.

Zusammenfassung. Glaubwürdigkeitsurteile zeichnen sich dadurch besonders aus, dass es in ihrem Rahmen auch um moralische Wertungen geht. Im Schulkontext können sie zudem bedeutende Konsequenzen nach sich ziehen

– im Rahmen der Erziehung die Ahndung unmoralischen Verhaltens und im Rahmen der Lehrer-Schülerbeziehungen etwaige Falschbeschuldigungen.

1.3 Täuschung im Schulkontext

Im Folgenden wird zunächst auf die subjektiv wahrgenommene Relevanz von Glaubwürdigkeitsurteilen in verschiedenen Schulalltagssituationen, die (verbale) Täuschung nach sich ziehen, eingegangen und anschließend die Häufigkeit verschiedener Täuschungssituationen dargestellt.

1.3.1 Subjektive Häufigkeit von Täuschungssituationen im Schulkontext

Es gibt verschiedene Situationen im Schulalltag, die als Täuschungssituationen eingestuft werden können, und in denen Lehrer die Einschätzung der Glaubwürdigkeit von Schülern für relevant halten. Dabei können Täuschungssituationen als Situationen verstanden werden, (1) in denen Schüler eine Handlung vollziehen, die – um Strafe zu vermeiden – eine Lüge bzw. verbale Täuschung nach sich zieht, und (2) in denen die Lehrkraft die Glaubwürdigkeit von Schülern einschätzen muss (vgl. Marksteiner, Reinhard, et al., 2013). Marksteiner, Reinhard, et al. untersuchten, in welchen Schulalltagssituationen es Lehrkräften relevant erscheint, die Glaubwürdigkeit von Schülern einzuschätzen (Studie 1), und wie diese Situationen von Lehrkräften wahrgenommen werden (Studie 1 und 2). Die Ergebnisse der ersten Studie deuten darauf hin, dass Lehrkräfte Schüler tatsächlich interviewten (bzw. theoretisch interviewen würden), wenn es zur Nutzung unerlaubter Hilfsmittel, zu aggressivem Verhalten gegenüber Mitschülern, Diebstahl, unerlaubter Abwesenheit, Bullying und Vandalismus kommt. Die befragten Lehrkräfte

gaben interessanterweise an, dass diese Situationen häufiger bei Kollegen als bei ihnen selbst auftreten. Sowohl in Studie 1 als auch in Studie 2 wurde das Nutzen unerlaubter Hilfsmittel, die Abwesenheit ohne Erlaubnis und Bullying (über einen längeren Zeitraum) als am häufigsten im Schulalltag auftretend eingestuft. Zu entdecken, ob ein Schüler die Wahrheit sagt oder lügt, wurde bei Situationen mit schwerwiegenden Konsequenzen (z.B. physische und psychische Beeinträchtigung bei Bullying) als am wichtigsten eingeschätzt. Somit existiert eine Reihe von Situationen, in denen Lehrkräfte die Einschätzung der Glaubwürdigkeit ihrer Schüler für relevant halten. Im Allgemeinen halten Lehrkräfte es allerdings für wichtiger, einen Schüler nicht fälschlicherweise eines Fehlverhaltens zu beschuldigen, als einen Lügner zu entlarven.

1.3.2 Häufigkeit von Täuschungssituationen im Schulkontext

Täuschungsversuche, z.B. die Nutzung eines Spickzettels, das Abschreiben vom Sitznachbarn oder Plagiiere, werden im Schulalltag sowohl von Lehrkräften als auch von Schülern als ernsthaftes Problem angesehen (Evans & Craig, 1990). Diese Wahrnehmung wird durch Befunde zur Häufigkeit von Täuschung im Schulalltag unterstützt (vgl. Davis, Grover, Becker, & McGregor, 1992; Franklyn-Stokes & Newstead, 1995). So gaben mehr als 75% der Schüler und Studenten an High Schools bzw. Colleges in den USA an, schon mindestens einmal während ihrer Schulzeit getäuscht zu haben, wobei weibliche Befragte durchgehend geringere Täuschungsraten berichteten als männliche (Davis et al., 1992). Franklyn-Stokes und Newstead (1995) berichten zudem, dass mehr als 60% der von ihnen befragten Schüler und Studierenden angaben,

Hausarbeiten abgeschrieben, Plagiate eingereicht bzw. Forschungsdaten verändert oder erfunden zu haben. Auch hier deuten die Befunde darauf hin, dass männliche Befragte häufiger täuschten als weibliche Befragte. Obwohl ältere Befragte Täuschungen als ernsthafteres Problem ansehen als jüngere Befragte, gibt es keinen Effekt des Alters in Bezug auf die angegebene Täuschungshäufigkeit. Die Aufgabe der Lehrkraft besteht nun darin, Täuschungsversuche aufzudecken, da nur so eine faire Bewertung der Schüler und eine angemessene Förderung und Gestaltung des Unterrichts möglich ist.

1.4 Erkennen von Lüge und Wahrheit

Nachfolgend soll ein Überblick über die Forschung zur Erkennungsleistung von wahren und erfundenen Aussagen gegeben werden. Dabei wird auch auf die Unterschiede zwischen sog. Experten und Laien sowie auf Urteilsverzerrungen eingegangen.

1.4.1 Durchschnittliche Trefferquote

In den meisten Untersuchungen in der Lügenforschung werden Probanden (*Empfänger* einer Nachricht oder *Rezipienten*) kurze Videosequenzen von Stimuluspersonen (*Sender* einer Nachricht) vorgelegt und sie werden gebeten einzuschätzen, ob die Sender die Wahrheit sagen oder lügen (vgl. C. F. Bond & DePaulo, 2006; Vrij, 2008). Meistens (vgl. C. F. Bond & DePaulo, 2006; Vrij, 2008) berichtet eine Hälfte der Sender die Wahrheit, während die andere Hälfte lügt. Sofern also die Empfänger in Bezug auf die Einschätzung, ob die Nachricht eines Senders wahr oder erfunden ist, raten

würden (Zufallsniveau²), würden sie bei den tatsächlich wahren Aussagen eine Trefferquote von 50%, bei den tatsächlich erfundenen Aussagen eine Trefferquote von 50% und zudem eine Gesamttrefferquote von ebenfalls 50% erzielen (Vrij, 2008, S. 147). Bisherige Befunde deuten darauf hin, dass sich die durchschnittliche *Trefferquote* beim Erkennen von wahren und erfundenen Aussagen mit 54% auf dem Zufallsniveau bewegt (C. F. Bond & DePaulo, 2006).

1.4.2 *Trefferquote von Laien und Experten*

In vielen der bisher durchgeführten Studien (für einen Überblick siehe Vrij, 2008, Anhang 6.1, S. 187) waren Empfänger der Nachricht *Studierende* oder sog. *Laien*, d.h. Personen, deren berufliche Tätigkeit es nicht beinhaltet, zwischen wahren und erfundenen Aussagen zu differenzieren. Zudem gibt es eine Reihe von Studien, bei denen die Empfänger einer Nachricht „professionelle catchers“ (vgl. Vrij, 2008; professionelle Lügendetektoren, eigene Übersetzung) oder *Experten* waren, d.h. Personen, die im Rahmen ihrer beruflichen Tätigkeit Glaubwürdigkeitseinschätzungen vornehmen. Zu diesen können laut Vrij (2008; für einen Überblick siehe Tabelle 6.3, S. 161) neben Mitarbeitern des CIA (Ekman, O'Sullivan, & Frank, 1999) und insbesondere Polizeibeamten (z.B. Garrido, Masip, & Herrero, 2004; Mann, Vrij, & Bull, 2004) auch Lehrkräfte (Reinhard, Dickhäuser, et al., 2011; Vrij et al., 2006) gezählt werden.

² Unter Zufallsniveau wird die Ratewahrscheinlichkeit verstanden, wenn 50% wahre und 50% erfundene Aussagen zur Beurteilung vorgelegt werden (vgl. Vrij, 2008, S. 141).

Obwohl es im pädagogischen Bereich von großer Bedeutung ist, wie akkurat die Glaubwürdigkeit von Schülern eingestuft wird, gibt es nur wenige Studien, bei denen Lehrkräfte als Empfänger fungierten. In den mir insoweit bekannten Studien zeigte sich, dass die durchschnittlich erzielte Trefferquote ebenfalls nicht höher lag als die Trefferquote durch Raten (Reinhard, Dickhäuser, et al., 2011; Vrij et al., 2006). Somit scheinen sich Lehrer beim Erkennen von Lüge und Wahrheit nicht wesentlich von Laien oder anderen Experten zu unterscheiden. In einer weiteren Studie von Reinhard, Dickhäuser, et al. (2011) wurden neben Lehrkräften zudem auch *Referendare* und *Lehramtsstudierende* gebeten, Aussagen von angeblichen Schülern einzuschätzen. Die auf Video aufgezeichneten Aussagen zeigten jeweils einen Schüler, der von einem angeblichen Lehrer zu einem vom Schüler bearbeiteten Test befragt wurde. Alle Befragten hatten in dem aufgezeichneten Interview rückgemeldet bekommen, dass sie ein sehr gutes Testergebnis erzielt hatten, und sie wurden gefragt, ob sie Hilfsmittel verwendet hätten. Eine Hälfte der Befragten hatte bei der Bearbeitung des Tests tatsächlich von einem Lösungsblatt abgeschrieben, die andere hatte den Test hingegen ohne weitere Hilfsmittel bearbeitet. Somit waren die Aussagen derjenigen Probanden, die tatsächlich das Lösungsblatt verwendet hatten, erfunden und die Aussagen derjenigen, die keine Hilfsmittel verwendet hatten, wahr. Hinsichtlich der durchschnittlichen Trefferquote unterschieden sich die Lehrkräfte nicht von den *angehenden* Lehrern (Referendare und Lehramtsstudierende). In einer Studie von Vrij et al. (2006) wurden zudem neben Lehrkräften auch Sozialarbeiter, Polizeibeamte und Studierende gebeten, Aussagen von 5- bis 6-

jährigen Kindern, Jugendlichen und Erwachsenen einzustufen. Hier war die Trefferquote für alle Berufsgruppen (d.h. für Lehrkräfte, Sozialarbeiter, Polizeibeamte und Studierende) mit 60% höher als die Ratewahrscheinlichkeit; es zeigten sich jedoch keine Effekte der Berufsgruppen auf die Trefferquote.

1.4.3 *Trefferquote und Urteilsverzerrungen*

Bei vielen Studien (für einen Überblick siehe Vrij, 2008, Anhang 6.1, S. 187 f.), bei denen Studierende oder Laien Aussagen einschätzten, lag die Trefferquote beim Erkennen von *wahren* Aussagen mit durchschnittlich 63.4% höher als die Trefferquote beim Erkennen von *erfundenen* Aussagen mit durchschnittlich 48.15% (Vrij, 2008, S. 148). Dieser Effekt wird auch als *Wahrheitseffekt* (veracity effect, eigene Übersetzung, Levine, Park, & McCornack, 1999) bezeichnet. Dagegen scheinen Experten, z.B. Polizeibeamte, durchweg besser im Erkennen von erfundenen im Vergleich zu wahren Aussagen zu sein (vgl. Vrij, 2008). Diese Befunde werden häufig dadurch erklärt, dass Versuchsteilnehmer einer *Urteilsverzerrung* (C. F. Bond & DePaulo, 2006; Vrij, 2008) unterliegen. Dabei kann zwischen zwei Arten der Urteilsverzerrung unterschieden werden, der so genannten *Wahrheitsverzerrung* (truth bias, eigene Übersetzung), d.h. der Tendenz, eine Aussage eher als wahr einzustufen, und der *Lügenverzerrung* (lie bias, eigene Übersetzung), d.h. der Tendenz, Aussagen eher als erfunden einzustufen (Vrij, 2008, S. 148). Eine Wahrheitsverzerrung konnte dagegen bei Lehrkräften, Sozialarbeitern und Laien gezeigt werden (C. F. Bond & DePaulo, 2006; Reinhard, Dickhäuser, et al., 2011; Vrij, 2008; Vrij et al., 2006). Eine Lügenverzerrung konnte bisher für Gefängnisinsassen (z.B. G. D. Bond, Malloy,

Thompson, Arias, & Nunn, 2004) und Polizeibeamte (Masip, Alonso, Garrido, & Anton, 2005; Meissner & Kassin, 2002) nachgewiesen werden. Des Weiteren zeigte sich, dass die Lügenverzerrung bei Polizeibeamten stärker wird, je länger sie im Polizeidienst tätig sind (Masip et al., 2005; Meissner & Kassin, 2002). In der bereits dargestellten Studie von Reinhard, Dickhäuser, et al. (2011) zeigte sich, dass Lehrkräfte im Vergleich zu Lehramtsstudierenden eine statistisch signifikant stärkere Wahrheitsverzerrung zeigen. Referendare liegen in Bezug auf die Wahrheitsverzerrung zwischen Lehrkräften und Lehramtsstudierenden und unterscheiden sich von beiden Gruppen nicht signifikant. Diese Befunde und auch die Befunde von Masip und Kollegen (2005) sowie Meissner und Kassin (2002) deuten darauf hin, dass sich bei den Expertengruppen „Polizeibeamter“ und „Lehrkraft“ mit Eintritt in den beruflichen Dienst die Tendenz, Aussagen eher als erfunden bzw. wahr einzustufen, verstärkt. Masip und Kollegen (2005) nehmen an, dass diese zunehmende Lügenverzerrung bei Polizeibeamten mit Eintritt in den Polizeidienst durch Sozialisation innerhalb der Polizeitruppe zustande kommt. Reinhard, Dickhäuser, et al. (2011) gehen davon aus, dass die spezifische Rolle, die Lehrkräfte in Interaktionen mit Schülern einnehmen, eine Erklärung für die zunehmende Wahrheitsverzerrung ist. So argumentieren sie, dass Lehrer-Schüler-Beziehungen negativ beeinflusst werden, wenn Lehrer zu häufig (unbegründet) ihre Schüler des Lügens beschuldigen und sie z.B. vor der gesamten Klasse fragen, ob sie einen Spickzettel verwendet haben. Gestützt wird dieses Argument auch durch die Befunde von Marksteiner, Reinhard, et al. (2013), die zeigen, dass es Lehrkräfte für wichtiger erachten, einen Schüler nicht fälschlicherweise eines

Fehlverhaltens zu beschuldigen, als jedes Fehlverhalten aufzudecken.

Marksteiner und Reinhard (2013) argumentieren, dass derartige

Urteilsverzerrungen durch *Urteilsziele* zustande kommen könnten. So nehmen sie an, dass Polizeibeamte das Ziel haben, jeden Lügner zu entlarven, während Lehrkräfte das Ziel haben, keine Falschbeschuldigungen vorzunehmen. Das Ziel, jeden Lügner zu entlarven, kann dann wiederum zu einer Lügenverzerrung führen; wohingegen das Ziel, keine Person fälschlicherweise zu beschuldigen, zu einer Wahrheitsverzerrung führen sollte.

Zusammenfassung. Dass es bei der Beurteilung der Glaubwürdigkeit (von Schülern) zu Verzerrungen und nachfolgend zu einer höheren Trefferquote bei wahren oder erfundenen Aussagen kommen kann, zeigen die oben beschriebenen Studien. Noch nicht erforscht wurde jedoch bislang, wie sich diese (durch Urteilsziele hervorgerufene) Urteilsverzerrungen auf den *Prozess* der Glaubwürdigkeitsbeurteilung auswirken. In der vorliegenden Arbeit soll daher der Einfluss von Urteilszielen auf den Prozess der Glaubwürdigkeitsbeurteilung und deren Akkuratheit genauer betrachtet werden.

1.5 Soziale Informationsverarbeitung

Das nachfolgende Kapitel dient zum einen dazu, die Modi der sozialen Informationsverarbeitung und deren Determinanten zu beleuchten, und zum anderen, Forschung zu Glaubwürdigkeitsurteilen als Prozess der sozialen Informationsverarbeitung näher zu betrachten. Daher wird in einem ersten Schritt auf Verarbeitungsmodi in einem Zwei-Prozess-Modell der sozialen Informationsverarbeitung, dem Heuristic-Systematic Model (HSM) von Chen

und Chaiken (1999), auf Determinanten dieser Modi und auf die Erfassung des Verarbeitungsmodus eingegangen. In einem zweiten Schritt wird ein Überblick über bisherige prozessorientierte Forschung zu Glaubwürdigkeitsurteilen gegeben.

1.5.1 *Modi der Informationsverarbeitung*

Verarbeitungsmodi im HSM. Es wurden bereits verschiedene Modelle der Informationsverarbeitung aufgestellt (vgl. Chaiken & Trope, 1999), wobei das HSM von Chen und Chaiken (1999) ein bekanntes und wissenschaftlich gut fundiertes Zwei-Prozess-Modell der Informationsverarbeitung darstellt. Das HSM macht im Vergleich zu anderen Modellen der Informationsverarbeitung (z.B. Elaboration-Likelihood-Model [ELM] von Cacioppo & Petty, 1984) klare Annahmen in Bezug auf Verzerrung bei der sozialen Informationsverarbeitung, die für die vorliegende Arbeit von besonderem Interesse und auch empirisch gut belegt sind (vgl. Chaiken & Trope, 1999).³ Im HSM werden zwei Verarbeitungsmodi der Informationsverarbeitung unterschieden, der *systematische* und der *heuristische* Modus. Wird *systematisch* verarbeitet, werden urteilsrelevante Informationen (relativ) analytisch und umfassend verarbeitet. Eine systematische Verarbeitung erfordert neben kognitiver

³ Im ELM (vgl. Chaiken & Trope, 1999) werden durchaus Annahmen über verzerrte Informationsverarbeitung getroffen (vgl. Postulat 3 zu multiplen Rollen und Postulat 5 zu verzerrter Verarbeitung), z.B. wird kurz erläutert, dass Stimmung die Verarbeitung beeinflusst. Das HSM macht jedoch spezifischere Annahmen darüber, wie verzerrte Informationsverarbeitung in Abhängigkeit des Verarbeitungsmodus abläuft (vgl. verzerrte heuristische und systematische Informationsverarbeitung).

Fähigkeit des Urteilenden auch dessen Motivation und Verarbeitungskapazität. Für die Bildung von Urteilen auf systematische Weise wird insbesondere der Nachrichteninhalt (z.B. Details) verwendet (Reinhard & Sporer, 2008, 2010). *Heuristische* Verarbeitung führt dagegen dazu, dass sowohl „Heuristiken“⁴, d.h. Daumenregeln (z.B. „Schüler sind unreif“), aktiviert und dann zur Urteilsbildung herangezogen werden, als auch urteilsrelevante Hinweisreize (z.B. die Länge einer Nachricht eines Senders) und nonverbale Informationen bzw. Quellenmerkmale (z.B. die Bewegung der Hände des Senders, Reinhard & Sporer, 2008, 2010). Im Gegensatz zur systematischen Verarbeitung sind bei heuristischer Verarbeitung lediglich minimale kognitive Leistungen erforderlich. Heuristische Verarbeitung wird durch so genannte sozial-kognitive Prinzipien bestimmt, d.h. die urteilsrelevanten Heuristiken müssen *verfügbar*, *zugänglich* und *anwendbar* sein. Dies würde bedeuten, dass die Heuristik „Schüler sind unreif“ im Gedächtnis gespeichert bzw. erlernt worden sein muss (Verfügbarkeit), in einer gegebenen Situation abrufbar sein muss (Zugänglichkeit) und zudem relevant für das Urteil sein muss (Anwendbarkeit).

⁴ „Heuristik“ ist hier gleichbedeutend mit „Daumenregel“ (vgl. Chen & Chaiken, 1999, S. 74). Auch in der Statistik wird der Begriff „Heuristik“ als „Daumenregel“ verstanden und wird dem Algorithmus entgegengesetzt (Bortz & Döring, 2006, S. 353). Im Vergleich zum Algorithmus, der Lösungsschritte bei einem Problem genau definiert und bei korrekter Befolgung mit Sicherheit das angestrebte Resultat erzielt (z.B. Algorithmus im Sinne einer Rechenregel), gibt eine Heuristik nur die ungefähre Richtung an und garantiert den Erfolg nicht (z.B. Heuristik im Sinne von Strategien der Partnersuche). Im Alltag sind Heuristiken von besonderer Bedeutung, da häufig in komplexen Situationen keine geeigneten Algorithmen existieren.

Determinanten des Verarbeitungsmodus. Eine Vielzahl an Arbeiten hat untersucht, durch welche Faktoren bestimmt wird, ob Informationen intensiv und mit viel kognitivem Aufwand oder oberflächlich und mit wenig kognitivem Aufwand bzw. systematisch oder heuristisch verarbeitet werden (vgl. Chaiken & Trope, 1999). Standardmäßig werden Informationen heuristisch und erst unter bestimmten Bedingungen systematisch verarbeitet. Chaiken, Liberman und Eagly (1989) weisen darauf hin, dass *persönliche Relevanz* (personal relevance; z.B. Petty, Goldman, & Cacioppo, 1981), *Aufgabenwichtigkeit* (task importance; z.B. Chaiken & Maheswaran, 1994; Maheswaran & Chaiken, 1991) und *interesseninkonsistente Informationen* (Giner-Sorolila & Chaiken, 1997) Faktoren sind, welche die systematische Verarbeitung von Informationen wahrscheinlicher machen. Bohner, Rank, Reinhard, Einwiller und Erb (1998) zählen neben Aufgabenwichtigkeit und persönlicher Relevanz auch *Beteiligung* (involvement; z.B. Leippe & Elkin, 1987), *wahrgenommene Konsequenzen* (perceived consequences; Chaiken, 1980), *persönliche Verantwortung* (personal responsibility; z.B. McAllister, Mitchell, & Beach, 1979) und *Interesse* zu den Determinanten des Verarbeitungsmodus. Wood (2000) nennt in ihrem Überblicksartikel zu Einstellungsänderung ebenfalls verschiedene Faktoren, die systematische, aufmerksame Verarbeitung erhöhen: *Rahmung* (framing; Smith & Petty, 1996), *Selbstbezug* (self-reference; Moreno & Mayer, 2000; Symons & Johnson, 1997; Turco, 1996), *symbolische Redewendungen* (token phrases; Garst & Bodenhausen, 1996), *feste, zugängliche Einstellungen* (strong, accessible attitudes; Fabrigar, Priester, Petty, & Wegener, 1998), *ambivalente Einstellungen* (ambivalent attitudes; Maio, Bell, & Esses, 1996),

Kognitionsbedürfnis (Need for Cognition; Cacioppo & Petty, 1984; Reinhard, 2010), *Urteilssicherheit* (judgment confidence; Bohner et al., 1998) und *zugängliche, wenig komplexe Sprache* (accessible, not overly complex language; Hafer, Reynolds, & Obertynski, 1996). Studien aus dem Bereich der Glaubwürdigkeitsforschung manipulierten den Verarbeitungsmodus durch *situative Vertrautheit* (Reinhard, Sporer, Scharmach, & Marksteiner, 2011) oder durch den *Zeitpunkt des Urteils* (decision moment; Masip, Garrido, & Herrero, 2006; Masip, Garrido, & Herrero, 2009).⁵

Erfassung heuristischer und systematischer Verarbeitung. Im vorangegangenen Abschnitt wurde dargelegt, unter welchen Umständen eine heuristische oder systematische Verarbeitung von Informationen auftreten kann. Nun soll erörtert werden, wie empirisch überprüft werden kann, ob der Urteilende heuristisch oder systematisch Informationen verarbeitet. In einigen Arbeiten (z.B. Chaiken, 1980; Chaiken & Maheswaran, 1994; Maheswaran & Chaiken, 1991) wurde dazu gezielt der Inhalt einer Nachricht variiert und als Ausdruck des Verarbeitungsmodus die Einstellungsänderung gegenüber einem Objekt (z.B. Chaiken & Maheswaran, 1994) oder zu einem Thema (z.B. Chaiken, 1980) und die Anzahl der erinnerten Argumente und Gedanken erfasst. Chaiken (1980) manipulierte den Verarbeitungsmodus, indem sie die wahrgenommenen Konsequenzen (perceived consequences) der Probanden, die alle Studierende waren, variierte: *hohe* Konsequenzen (high consequences) wurden induziert, indem den Probanden mitgeteilt wurde, dass sie zunächst ihre Meinung zu

⁵ Diese Aufzählung soll einen kurzen Überblick über verschiedene Determinanten des Verarbeitungsmodus geben und keineswegs den Anspruch haben, vollständig zu sein.

einem bestimmten Thema (Thema A) äußern sollten und anschließend über dieses Thema (Thema A) mit anderen Probanden diskutieren würden; *niedrige* Konsequenzen (low consequences) wurden induziert, indem den Probanden mitgeteilt wurde, dass sie zunächst ihre Meinung zu einem bestimmten Thema (Thema A) äußern sollten, anschließend aber über ein anderes Thema (Thema B) mit anderen Probanden diskutieren sollten. Anschließend wurde allen Probanden ein Interview vorgelegt, in dem eine Person (Interviewte) zu Thema A befragt wurde. Die Beliebtheit des Interviewten wurde variiert, indem der Interviewte entweder als studentenfreundlich oder studentenfeindlich dargestellt wurde. Dabei wurde angenommen, dass die Beliebtheit des Interviewten ein heuristischer Hinweisreiz ist, der bei heuristischer Verarbeitung in der Konsequenz zu einer Einstellungsänderung führen sollte, bei systematischer Verarbeitung aber keine Relevanz haben sollte. Neben der Beliebtheit des Interviewten wurde der Inhalt des Interviews variiert und der Interviewte nannte entweder zwei Argumente (wenig überzeugende Argumentation) oder sechs (überzeugende Argumentation), die gegen die Meinung der Probanden sprach⁶. Es wurde angenommen, dass die Anzahl der Argumente die Einstellungsänderung bei systematischer Verarbeitung beeinflussen, bei heuristischer Verarbeitung jedoch keinen Einfluss auf die Einstellungsänderung haben sollten. So wurde davon ausgegangen, dass die Probanden bei systematischer Verarbeitung eine positivere Meinung zu Thema A haben sollten, wenn die Argumentation des Interviewten überzeugend im

⁶ Die Meinung über Thema A wurde vorgetestet und war im Durchschnitt negativ, die Argumente des Interviewten waren dagegen positiv.

Vergleich zu wenig überzeugend war. Nach Lesen des Interviews wurden die Probanden zudem gebeten, die im Interview genannten Argumente wiederzugeben. Die Ergebnisse deuten darauf hin, dass sich bei *hohen* Konsequenzen die Einstellung der Probanden stärker änderte, wenn der Interviewte sechs im Vergleich zu zwei einstellungskonträren Argumenten nannte. Die Beliebtheit des Interviewten hatte dabei keinen Einfluss auf die Einstellungsänderung, was auf systematische Verarbeitung schließen lässt. Bei *niedrigen* Konsequenzen dagegen hatte die Anzahl der Argumente keinen Einfluss auf die Einstellungsänderung, die Beliebtheit des Interviewten jedoch schon: Eine größere Einstellungsänderung wurde durch einen beliebten im Vergleich zu einem unbeliebten Interviewten erzielt, was auf eine heuristische Verarbeitung schließen lässt. Des Weiteren wurden bei hohen im Vergleich zu niedrigen Konsequenzen mehr Argumente erinnert, wobei dieser Unterschied stärker war, wenn sechs im Vergleich zu zwei Argumenten vom Interviewten genannt wurden. Die Anzahl erinnelter Argumente wurde als Hinweis auf eine systematische oder heuristische Informationsverarbeitung angesehen. Es lässt sich folglich empirisch feststellen, welcher Verarbeitungsmodus vorliegt.

Zusammenfassung. Zusammenfassend lässt sich sagen, dass die Verarbeitung von Informationen als ein durch verschiedene Determinanten bestimmter Prozess verstanden werden kann und dass in Abhängigkeit vom Verarbeitungsmodus (systematisch vs. heuristisch) unterschiedliche Informationen (z.B. Nachrichteninhalt, Argumente, Heuristiken, Hinweisreize, Handbewegungen) zur Urteilsbildung verwendet werden. Welche Verarbeitungsstrategie verfolgt wird, wird zum einen indirekt über die

Einstellung und zum anderen über Gedanken und die Wiedergabe von Nachrichteninhalten erfasst.

1.5.2 Informationsverarbeitung und Glaubwürdigkeitsurteile

Glaubwürdigkeitsbeurteilung als Informationsverarbeitungsprozess.

Wie bereits oben erwähnt, kann die Beurteilung der Glaubwürdigkeit einer Person als Prozess der sozialen Informationsverarbeitung verstanden werden (Reinhard & Sporer, 2008, 2010). Bei den meisten Arbeiten zur Untersuchung von Urteilsprozessen (vgl. Chaiken & Trope, 1999) lag der Fokus jedoch insbesondere auf der *Einstellung* zu einem Thema (z.B. Bohnet et al., 1998; Cacioppo, Kao, Petty, & Rodriguez, 1986; Cacioppo, Petty, & Morris, 1983; Chaiken, 1980) oder zu einem Produkt (z.B. Chaiken & Maheswaran, 1994; Maheswaran & Chaiken, 1991) und nur bei wenigen Arbeiten auf der *Einschätzung einer Person* (z.B. Reinhard & Sporer, 2008, 2010). Um aber zu verstehen, warum Urteilende ein bestimmtes Urteil über eine Person, z.B. bezüglich der Glaubwürdigkeit, abgeben, ist es hilfreich, den zugrunde liegenden Prozess der Urteilsbildung zu verstehen und mögliche Motivationen des Urteilers zu kennen. Somit ist es wichtig zu erforschen, *wie* derartige Urteile entstehen. Ist der Prozess der Urteilsbildung bekannt, sollte auch verständlich werden, warum die Akkuratheit der meisten Glaubwürdigkeitsurteile nicht höher ist als die Ratewahrscheinlichkeit (vgl. Kapitel 1.4 und C. F. Bond & DePaulo, 2006). Dieses Verständnis könnte auch bei der Suche danach hilfreich sein, wie die Akkuratheit von Glaubwürdigkeitsurteilen erhöht werden kann. Bisher gibt es jedoch nur wenige Arbeiten, die sich mit dem Prozess der Glaubwürdigkeitsbeurteilung auseinandergesetzt haben (Reinhard, 2010;

Reinhard & Sporer, 2008, 2010). Reinhard und Sporer konnten z.B. zeigen, dass Aufgabeneinbindung (task involvement) und kognitive Kapazität bestimmen, welche Informationen zur Beurteilung der Glaubwürdigkeit einer Person herangezogen werden (Reinhard & Sporer, 2008, 2010). Die Aufgabeneinbindung wurde variiert, indem den Probanden entweder berichtet wurde, dass die Ergebnisse der Studie, an der sie teilnahmen, äußerst wichtig für die zukünftige psychologische Forschung sind, oder dass die Daten lediglich für einen Einführungskurs im Bereich Methoden genutzt werden. Es zeigte sich, dass Personen mit starker Aufgabeneinbindung und ausreichend kognitiver Kapazität eher verbale Informationen bzw. den Nachrichteninhalt (z.B. Details) zur Urteilsbildung nutzten. Dagegen verwendeten Personen mit schwacher Aufgabeneinbindung und wenig kognitiver Kapazität nur nonverbale Informationen bzw. Quellenmerkmale (z.B. die Hand- und Fußbewegungen des Senders).

Aber nicht nur situationsbezogene Faktoren, wie Aufgabeneinbindung und kognitive Kapazität, sondern auch personenbezogene Faktoren wie die Persönlichkeitsvariable *Need for Cognition* (NFC; Cacioppo & Petty, 1982) beeinflussen die Nutzung unterschiedlicher Informationen im Rahmen von Glaubwürdigkeitsurteilen. NFC kann als Kognitionsbedürfnis (vgl. Aronson, Wilson, & Akert, 2008, deutsche Übersetzung) verstanden werden, d.h. als Spaß und Freude am Nachdenken. Die Befunde von Reinhard (2010) deuten darauf hin, dass Personen mit hohem NFC besser im Erkennen von Lüge und Wahrheit sind, was dadurch erklärt wird, dass Personen mit hohem NFC neben nonverbalen Informationen auch verbale Informationen bei ihrer Urteilsbildung

nutzen. Die Ergebnisse von Reinhard, Sporer, et al. (2011) stützen diese Befunde, indem sie zeigen, dass Personen, die sich aufgrund ihrer erhöhten situativen Vertrautheit stärker auf verbale Informationen stützen, Lüge und Wahrheit besser unterscheiden können. Dies kann dadurch erklärt werden, dass mehr verbale Indikatoren (z.B. Detailreichtum) im Vergleich zu nonverbalen Indikatoren als valide Indikatoren angesehen werden können (C. F. Bond & DePaulo, 2006), d.h. als Indikatoren, die tatsächlich einen Hinweis darauf geben, ob eine Person lügt oder die Wahrheit sagt.

Glaubwürdigkeitsurteile im pädagogischen Bereich. Obwohl das akkurate Fällen von Glaubwürdigkeitsurteilen im pädagogischen Kontext eine bedeutende Rolle spielt (vgl. Kapitel 1.3), gibt es nur wenige Arbeiten, die sich mit Glaubwürdigkeitsurteilen im pädagogischen Bereich befassen haben (Dickhäuser, Reinhard, & Marksteiner, 2012; Marksteiner, Dickhäuser, & Reinhard, 2012; Marksteiner, Reinhard, et al., 2012; Marksteiner, Reinhard, et al., 2013; Reinhard, Dickhäuser, et al., 2011; Vrij et al., 2006). In allen Studien zeigte sich, dass Lehrer bzw. angehende Lehrer beim Erkennen von Lüge und Wahrheit nicht besser sind, als wenn sie geraten hätten. Damit sind Lehrer nicht besser als Laien (vgl. C. F. Bond & DePaulo, 2006). Obwohl die Erforschung des Prozesses der Glaubwürdigkeitsbeurteilung Hinweise auf die Ursache der geringen Trefferquoten liefert und Glaubwürdigkeitsbeurteilungen im pädagogischen Bereich bedeutsam sind (vgl. Kapitel 1.3), gibt es noch keine Arbeiten, die den *Prozess* von Glaubwürdigkeitsbeurteilungen im pädagogischen Bereich untersucht haben.

Zusammenfassung. Glaubwürdigkeitsurteile können somit als Informationsverarbeitungsprozesse verstanden werden, wobei sowohl situative als auch personenbezogene Variablen beeinflussen, welche Informationen (verbale vs. nonverbale Informationen) zur Urteilsbildung herangezogen werden. Es gibt Studien, die Glaubwürdigkeitsurteile im pädagogischen Kontext untersuchten, jedoch lassen sich keine Untersuchungen finden, die den *Prozess* der Glaubwürdigkeitsbeurteilung im *pädagogischen Bereich* erforschten. Die Erforschung des Prozesses kann Erkenntnisse liefern, die dann wiederum genutzt werden können, um die Trefferquote bei Glaubwürdigkeitsurteilen zu erhöhen.

1.6 Verzerrungen bei der Informationsverarbeitung

Das nachfolgende Kapitel dient zum einen dazu, die Rolle von Motivation und Zielen bei der Informationsverarbeitung näher zu betrachten, und zum anderen, die Bedeutsamkeit der Forschung zu (verzerrten) Glaubwürdigkeitsurteilen zu beleuchten. Dazu wird in einem ersten Schritt (Kapitel 1.6.1) auf die Darstellung zu Informationsverarbeitung und motivationale Verzerrungen bei Kunda (1990) und auf weitere Annahmen zum Einfluss von Motivation auf die Informationsverarbeitung im HSM von Chen und Chaiken (1999) eingegangen. Dabei wird versucht, Gemeinsamkeiten zwischen den Annahmen von Kunda sowie Chen und Chaiken aufzuzeigen. Im Anschluss wird die *verzerrte* systematische und heuristische Informationsverarbeitung ausführlich erläutert. In einem zweiten Schritt (Kapitel 1.6.2) wird ein Überblick über bisherige Forschung zu *verzerrten*

Glaubwürdigkeitsurteilen gegeben und drei zentrale Studien zu verzerrten Glaubwürdigkeitsurteilen werden ausführlich beschrieben.

1.6.1 Motivation, Ziele und verzerrte Informationsverarbeitung

Informationsverarbeitung als motivationaler Prozess. Kunda (1999)

unterscheidet bei der Informationsverarbeitung zwischen „Cold Cognition“ und „Hot Cognition“ (S. 211). Unter „Cold Cognition“ (kalter Denkprozess, eigene Übersetzung) werden relativ intellektuelle, informationsgesteuerte Denkprozesse zusammengefasst, unter „Hot Cognition“ (heißer Denkprozess, eigene Übersetzung) dagegen motivationale, affektgeladene Prozesse. Sowohl „kalte“ bzw. nicht-motivationale als auch „heiße“ bzw. motivationale Denkprozesse können verzerrt sein. Bei nicht-motivationalen Denkprozessen kann die Verarbeitung weiterer Informationen beispielsweise durch Erfahrungen und Wissen verzerrt werden, wohingegen bei motivationalen Denkprozessen z.B. die Motivation, sich selbst in einem positiven Licht zu sehen, die Informationsverarbeitung beeinflusst (S. 214). „Hot Cognition“ bezieht sich somit auf diejenigen mentalen Prozesse, die durch Wünsche und Gefühle angetrieben werden, also auf diejenigen Situationen, in denen Ziele und auch Emotionen das Urteil beeinflussen (Kunda, 1999, S. 211). Für die vorliegende Arbeit ist insbesondere Informationsverarbeitung im Sinne von „Hot Cognition“ bzw. Informationsverarbeitung als „heißer“, motivationaler Denkprozess von Bedeutung. In Bezug auf *Verhalten* kann *Motivation* verstanden werden als

„...die Energie in uns, die unser Verhalten aktiviert und in Richtung eines bestimmten Ziels lenkt“ (Phares & Chaplin, 1997, S. 434, eigene Übersetzung, zitiert nach Elliot & Fryer, 2008, S. 235).

In Bezug auf *Informationsverarbeitung* kann Motivation als etwas angesehen werden, das unsere Wahrnehmung aktiviert und in Richtung eines bestimmten *Urteilsziels* lenkt.

Ziele und Informationsverarbeitung. Kunda (1990) unterscheidet in ihrem viel zitierten Artikel über „Motivated Reasoning“ verschiedene Ziele, welche die Informationsverarbeitung beeinflussen. Dabei differenziert sie zwischen *Akkuratheitszielen* (accuracy goals) und *Richtungszielen* (directional goals). *Akkuratheitsziele* führen dazu, dass Personen motiviert sind, eine äußerst angemessene Schlussfolgerung zu ziehen bzw. ein möglichst akkurates und zutreffendes und damit unverzerrtes Urteil zu fällen. Bei der Verfolgung eines *Richtungsziels* sind Personen dagegen motiviert, eine bestimmte Schlussfolgerung zu ziehen bzw. zu einem spezifischen Urteil zu kommen, was zu Verzerrungen führen kann. Man kann also zwischen Zielen unterscheiden, die zu einer relativ unverzerrten Informationsverarbeitung führen (Akkuratheitsziele), und solchen, welche die Verzerrung der vorliegenden Informationen bewirken (Richtungsziele). Was unter einem *Ziel* verstanden werden kann, ist laut Kunda häufig abhängig vom Kontext, in dem dieses Konstrukt verwendet wird. Kontextunabhängig kann ein *Ziel* jedoch wie folgt definiert werden:

“A goal is what an individual is trying to accomplish; it is the object or aim of an action” (Locke, Saari, Shaw, & Latham, 1981, S. 126).

Ein Ziel ist also etwas, das ein Individuum zu erreichen versucht, das Objekt oder die Absicht einer Handlung. Wichtig ist hierbei, dass Ziele nicht immer bewusst sein müssen, sondern auch unbewusst unser Handeln und Denken beeinflussen können (vgl. Uleman, 1996, 1999). Wie Uleman in seinem Artikel über „When unconscious goals cloud our minds“ (Uleman, 1996) und in seinem Aufsatz zu spontanen und intentionalen Schlussfolgerungen bei der Eindrucksbildung (Uleman, 1999) beschreibt, sind Ziele nicht immer bewusst und können daher nicht direkt durch die Befragung von Probanden erfasst werden (vgl. Uleman, 1996, 1999). Nur Ziele, die bewusst sind, können damit direkt erfasst werden.

Motivation und Informationsverarbeitung. Im HSM (Chen & Chaiken, 1999) wird zwischen drei unterschiedlichen Arten der Motivation unterschieden, welche die Informationsverarbeitung beeinflussen können: *Akkuratheitsmotivation* (accuracy motivation), *Verteidigungsmotivation* (defense motivation) und *Eindrucksmotivation* (impression motivation). Während Personen, die aufgrund von *Akkuratheitsmotivation* handeln, urteilsrelevante Informationen relativ offen und unvoreingenommen verarbeiten, ist die Informationsverarbeitung bei Personen mit einer Verteidigungs- oder Eindrucksmotivation verzerrt. *Verteidigungsmotivation* bezieht sich dabei auf den Wunsch, Einstellungen und Überzeugungen zu haben, die mit den eigenen Interessen, den Wertvorstellungen, der sozialen Identität oder den persönlichen Eigenschaften übereinstimmen. Ziel des Verteidigungsmotivierten ist es, das eigene Selbstkonzept und die eigene Weltanschauung zu erhalten. Als Folge werden Informationen selektiv

verarbeitet, d.h. sie werden derart interpretiert und genutzt, dass sie das Erreichen der genannten Ziele ermöglichen. *Eindrucks motivation* bezieht sich auf den Wunsch, Einstellungen und Überzeugungen zu haben, die im Einklang mit aktuellen sozialen Zielen stehen. Dabei geht es insbesondere darum, bei einer sozialen Interaktion in einem gegebenen sozialen Kontext einen bestimmten Eindruck zu hinterlassen oder das Vorhandensein einer Einstellung zu vermitteln. Auch hier kann es zu Verzerrungen kommen. Es gibt also Arten von Motivation, die entweder zu einer relativ unverzerrten Verarbeitung von Informationen führen (Akkuratheitsmotivation) oder zur Verarbeitung von Informationen auf eine eher verzerrte Art und Weise (Verteidigungsmotivation, Eindrucks motivation).

Motivation und Ziele. Da in vielen Definitionen des Begriffs „Motivation“ das Konstrukt „Ziel“ einbezogen wird (Elliot & Fryer, 2008, S. 235), überschneiden sich die Annahmen des HSM von Chen und Chaiken (1999) und der Ansatz von Kunda (1990, 1999). Beiden ist gemein, dass sie einerseits von einer Motivation (Akkuratheitsmotivation) oder einem Ziel (Akkuratheitsziel) ausgehen, die bzw. das zu einer relativ unvoreingenommenen Informationsverarbeitung führt, und einer Motivation (Verteidigungs- und Eindrucks motivation) oder einem Ziel (Richtungsziel), durch die bzw. das die Verarbeitung von Informationen auf eine verzerrte Art und Weise stattfindet. Beide Ansätze unterscheiden also die relativ unverzerrte Verarbeitung von Informationen und die eher verzerrte Informationsverarbeitung.

Verzerrte heuristische und systematische Informationsverarbeitung.

Wie zuvor erläutert (vgl. Kapitel 1.5.1), werden im HSM zwei Verarbeitungsmodi

unterschieden: der kognitiv weniger aufwendige *heuristische* Verarbeitungsmodus, bei dem heuristische Hinweisreize und nonverbale Informationen bzw. Quellenmerkmale genutzt werden, und der kognitiv aufwendige *systematische* Verarbeitungsmodus, bei dem insbesondere verbale Informationen bzw. Nachrichteninhalte berücksichtigt werden. Ist ein Urteiler verteidigungs- oder eindrucksmotiviert, nimmt das HSM an, dass die Verarbeitung von Informationen verzerrt erfolgt. Genauer gesagt geht es von einer *verzerrten heuristischen* und einer *verzerrten systematischen* Verarbeitung aus. Chen und Chaiken (1999) nehmen an, dass verteidigungsmotivierte Personen die gleichen Heuristiken (z.B. Schüler sind unreif) wie akkuratheitsmotivierte Personen verwenden, jedoch auf eine verzerrte Art und Weise. Laut den Autoren bedeutet „verzerrte heuristische Verarbeitung“, dass Heuristiken mit urteilsrelevanten Implikationen mit hoher Wahrscheinlichkeit angewendet werden, wenn sie das Ziel des Verteidigungsmotivierten unterstützen. Unterstützen sie das Ziel jedoch nicht, so ist die Wahrscheinlichkeit groß, dass sie nicht angewendet werden und sogar abgewertet und/oder ignoriert werden. Die Autoren erläutern weiterhin, dass „verzerrte systematische Verarbeitung“ durch eine aufwendige, aber verzerrte Prüfung und Bewertung von urteilsrelevanten Informationen stattfindet. Dabei werden Informationen, die konsistent mit den eigenen Einstellungen und Interessen sind, positiver bewertet als interessen~~in~~konsistente Informationen.

Erfassung verzerrter systematischer und heuristischer

Informationsverarbeitung. Eine Untersuchung, die von zentraler Bedeutung für die vorliegende Arbeit ist, stammt von Giner-Sorolla und Chaiken (1997).

Die Autoren untersuchten den Einfluss der Verteidigungsmotivation auf den Verarbeitungsmodus. Die Verteidigungsmotivation wurde nicht experimentell induziert, sondern es wurden nur Studierende mit einem großen Interesse an einem bestimmten Thema als Probanden ausgewählt: Die Studierenden hatten entweder ein eindeutiges Interesse daran, dass die Abschlussprüfungen an ihrer Universität in Form von Aufsätzen erfolgen oder in Form von Multiple-Choice-Tests durchgeführt werden. Den Probanden wurde in einem ersten Schritt ein heuristischer Hinweisreiz präsentiert. Dabei handelte es sich angeblich um die Ergebnisse einer Voruntersuchung zum Thema „Prüfungsarten an Ihrer Universität“. Eine Hälfte der Studierenden erhielt das Ergebnis, dass die meisten Teilnehmer der Voruntersuchung Aufsätze als Prüfungsart bevorzugten; die andere Hälfte bekam mitgeteilt, dass die meisten Teilnehmer Multiple-Choice-Prüfungen präferierten. Anschließend sollte der Hinweisreiz hinsichtlich seiner Zuverlässigkeit beurteilt werden. In einem zweiten Schritt wurde den Probanden eine Sammlung von Kommentaren vorgelegt, die angeblich von den Probanden der Voruntersuchung stammten. Die Sammlung enthielt insgesamt sechs Kommentare: drei Kommentare, die Aufsätze als Abschlussprüfung als positiv darstellten, und drei Kommentare, die Multiple-Choice-Tests als Prüfungsmethode der Wahl darstellten. Anschließend wurden die Probanden gebeten, (1) ihre Einstellung zu der Art der Abschlussprüfung einzustufen, (2) Gedanken aufzulisten, die sie während des Lesens der Ergebnisse der Voruntersuchung und der Kommentarsammlung hatten, und (3) die Stärke der Argumente, die in den Kommentaren genannt wurden, einzuschätzen. Die Ergebnisse deuten darauf hin, dass das Interesse bzgl. der

Art der Abschlussarbeit die Verarbeitung des heuristischen Hinweisreizes verzerrte: Waren die Ergebnisse der Voruntersuchung interessen~~in~~konsistent, wurde der Hinweisreiz bzw. die Voruntersuchung als weniger zuverlässig eingestuft und negativer beurteilt als bei interessenkonsistenten Ergebnissen. Weiterhin zeigen die Befunde, dass heuristische Hinweise, die als Bedrohung der eigenen Interessen wahrgenommen werden (interessen~~in~~konsistente Hinweisreize), dazu führten, dass Personen nachfolgende Informationen auf eine verzerrte *systematische* Art und Weise verarbeiten. Eine verzerrte systematische Verarbeitung zeigte sich einerseits dadurch, dass insgesamt mehr Gedanken geäußert wurden, nachdem die Probanden einen interessen~~in~~konsistenten (vs. interessenkonsistenten) Hinweisreiz gelesen hatten. Andererseits wurden die in den Kommentaren genannten Argumente, welche die eigenen Interessen unterstützten, als überzeugender eingestuft, als Argumente, welche die eigenen Interessen bedrohten. Ein weiteres Zeichen für die verzerrte systematische Verarbeitung war, dass die geäußerten Gedanken den Einfluss des Interesses auf die endgültige Einstellung mediierte, sofern zuvor ein interessen~~in~~konsistenter Hinweisreiz präsentiert wurde. Bei einem interessenkonsistenten Hinweisreiz zeigte sich hingegen kein Mediationseffekt der spezifischen Gedanken.

Zusammenfassung. Motivation beeinflusst unsere Wahrnehmung und lenkt sie in Richtung eines bestimmten Urteils(ziels). Dabei kann zwischen Richtungszielen bzw. Eindrucks- und Verteidigungsmotivation und Akkuratheitszielen bzw. Akkuratheitsmotivation unterschieden werden, wobei erstere die Informationsverarbeitung verzerren können. Besteht z.B. eine

Verteidigungsmotivation, so kann es sowohl zu einer verzerrten systematischen als auch zu einer verzerrten heuristischen Informationsverarbeitung kommen. Die Art der Informationsverarbeitung (systematisch vs. heuristisch) kann dabei anhand der Anzahl der geäußerten Gedanken erfasst werden. Die Verzerrung dagegen kann anhand der Valenz der geäußerten Gedanken und anhand der Bewertung der in den Kommentaren genannten Argumente gemessen werden.

1.6.2 Verzerrte Informationsverarbeitung und Glaubwürdigkeitsurteile

Wie in den vorangegangenen Kapiteln beschrieben wurde, kann Informationsverarbeitung als zum einen durch Ziele (vgl. Kunda, 1990) bzw. Verteidigungsmotivation (Chen & Chaiken, 1999) und zum anderen durch Verarbeitungsmodi (vgl. Chen & Chaiken, 1999) beeinflusster Prozess verstanden werden. Die Beurteilung der Glaubwürdigkeit einer Person kann wiederum als Prozess der Informationsverarbeitung angesehen werden (vgl. Reinhard & Sporer, 2008, 2010). Wie ebenfalls bereits erläutert wurde, zeigen Probanden in Studien zum Erkennen von Lüge und Wahrheit in den meisten Fällen eine Wahrheitsverzerrung (C. F. Bond & DePaulo, 2006) oder – insbesondere bei polizeilichen Ermittlungsbeamten – eine Lügenverzerrung (Vrij, 2008). Trotz dieser Erkenntnisse gibt es bisher nur wenige Arbeiten, die sich mit dem Prozess der *verzerrten* Informationsverarbeitung bei Glaubwürdigkeitsbeurteilung befassen (Hubbell, Mitchell, & Gee, 2001; Hurst-Wagner & Oswald, 2011; Masip et al., 2009). Diese Studien sollen nun ausführlich dargestellt werden, da sie von zentraler Bedeutung für die vorliegende Arbeit sind.

Einfluss des Zeitpunkts der Misstrauensinduktion und der Ergebnisinvolviertheit auf die Urteilsverzerrung. Hubbell und Kollegen (2001) untersuchten in zwei Studien, wie sich der Zeitpunkt einer Misstrauensinduktion (timing of suspicion) und die Ergebnisinvolviertheit (outcome involvement) auf die Beurteilung der Glaubwürdigkeit einer Stimulusperson auswirken.

In Studie 1 wurde den Probanden (die alle Studierende einer bestimmten Universität waren) eine ca. 10-minütige Videosequenz vorgespielt, in der eine Stimulusperson bei der Beantwortung mehrerer Fragen zu sehen war. Der Hälfte der Probanden wurde *vor* dem Abspielen der Videosequenz mitgeteilt, dass man der Stimulusperson nicht trauen könne, da sie in der Vergangenheit gelogen habe (Prä-Misstrauensinduktion). Der anderen Hälfte der Probanden wurde dies mitgeteilt, *nachdem* ihnen die Videosequenz vorgespielt worden war (Post-Misstrauensinduktion). Ergebnisinvolvierteit (hoch vs. niedrig) wurde induziert, indem der einen Hälfte der Probanden mitgeteilt wurde, dass die Stimulusperson sich auf einen Stelle im administrativen Bereich ihrer Universität bewerben würde (hohe Ergebnisinvolviertheit). Die andere Hälfte der Probanden erhielt keine weiteren Informationen über die Stimulusperson (niedrige Ergebnisinvolviertheit). Während des Betrachtens der Videosequenz wurden alle Probanden gebeten, jede der Antworten der Stimulusperson hinsichtlich ihrer Glaubwürdigkeit zu beurteilen. Die Autorinnen gingen davon aus, dass eine Prä-Misstrauensinduktion negative Erwartungen hinsichtlich der Glaubwürdigkeit der Stimulusperson entstehen lassen würde. Diese negativen Erwartungen sollten dazu führen, dass die Probanden der Stimulusperson

gegenüber misstrauischer sind und in der Folge mehr nonverbale Verhaltensweisen als „verdächtig“ oder „suspekt“ enkodieren würden als die Probanden der Post-Misstrauensinduktion (z.B. sollten sie oft auftretende Handbewegungen der Stimulusperson als Nervosität interpretieren und die Stimulusperson somit als unglaubwürdig beurteilen). Diese Art der Enkodierung sollte wiederum zu einer Lügenverzerrung führen, d.h. die Aussagen der Stimulusperson sollten eher als erfunden eingestuft werden. Des Weiteren sollten die Probanden der Prä-Misstrauensinduktions-Bedingung insgesamt eine höhere Anzahl an Verhaltensweisen auflisten, die sie als verdächtig interpretiert hatten. Zudem wurde angenommen, dass diese Effekte auf die Lügenverzerrung und die Anzahl genannter „verdächtiger“ Verhaltensweisen bei Personen mit hoher Ergebnisinvolviertheit stärker sein sollte: Wenn der Proband von Anfang an misstrauisch war (aufgrund der Prä-Misstrauensinduktion) und die Aussagen gründlich prüfte (aufgrund der hohen Ergebnisinvolviertheit), sollte er mehr verdächtige Verhaltensweisen nennen und eine stärker ausgeprägte Lügenverzerrung aufweisen.

Die Befunde von Studie 1 zeigen indes, dass der Zeitpunkt der Misstrauensinduktion keinen Effekt auf die Lügenverzerrung hatte. Des Weiteren zeigte sich, dass – entgegen der Hypothesen – bei *niedriger* (und nicht wie angenommen bei hoher) Involviertheit die Stimulusperson als verdächtig wahrgenommen wurde, d.h. dass Probanden mit niedriger Ergebnisinvolviertheit einer stärkeren Lügenverzerrung unterlagen. Auch gab es keinen Effekt auf die Anzahl genannter „verdächtiger“ Hinweise.

In Studie 2 wurde ebenfalls angenommen, dass bei einer Prä-Misstrauensinduktion und einer hohen Ergebnisinvolviertheit die Stimulusperson als am stärksten verdächtig wahrgenommen würde und die Probanden eine ausgeprägte Lügenverzerrung aufweisen sollten. Des Weiteren wurde ein weiterer Faktor (Status der Aussage) eingeführt: Die Stimulusperson sagte entweder ausschließlich die Wahrheit oder log (in Studie 1 waren die Aussagen alle wahr).

Die Ergebnisse von Studie 2 zeigten wieder entgegen der Hypothesen, dass die Stimulusperson erstens bei hoher (vs. niedriger) Ergebnisinvolviertheit als *ehrlicher* wahrgenommen wurde. Bei einer Prä-Misstrauensinduktion wurde die Stimulusperson zweitens als unehrlicher wahrgenommen als bei einer Post-Misstrauensinduktion. Insgesamt zeigte sich jedoch keine ausgeprägte Lügenverzerrung; es lag eher eine weniger ausgeprägte Wahrheitsverzerrung vor. Interessanterweise wurde die Stimulusperson als ehrlicher wahrgenommen, wenn die Aussagen alle erfunden (vs. wahr) waren. Eine signifikante Dreifachinteraktion zwischen Misstrauensinduktion, Ergebnisinvolviertheit und Status der Aussagen deutete darauf hin, dass die Stimulusperson bei hoher Ergebnisinvolviertheit, Post-Misstrauensinduktion und erfundenen Aussagen als am ehrlichsten wahrgenommen wurde. Als am wenigsten ehrlich wurde sie dagegen bei hoher Ergebnisinvolviertheit, Prä-Misstrauensinduktion und wahren Aussagen beurteilt. Wider Erwarten zeigte sich zudem kein Effekt der Misstrauensinduktion auf die Anzahl genannter „verdächtiger“ Verhaltensweisen.

Ziel der Studie von Hubbell et al. (2001) war es, den Effekt des Zeitpunkts der Misstrauensinduktion und der Ergebnisinvolviertheit auf die Verzerrungen bei der Informationsverarbeitung und auf das Glaubwürdigkeitsurteil zu untersuchen. Beide Studien zeigen keine hypothesenkonformen Effekte, deuten jedoch darauf hin, dass bei hoher Involviertheit Stimuluspersonen als weniger glaubwürdig wahrgenommen werden als bei niedriger Involviertheit. Da es keinen Effekt auf die Anzahl genannter „verdächtiger“ Hinweise gab, konnte der zugrundeliegende Prozess nicht aufgedeckt werden. Aufgrund des Designs konnte zudem die Akkuratheit der Glaubwürdigkeitsbeurteilung nicht untersucht werden.

Einfluss der Gewichtung von Fehlervermeidung auf die Urteilsverzerrung. Hurst-Wagner und Oswald (2011) untersuchten anhand von zwei Studien, ob Personen, die besorgt waren, jemanden fälschlicherweise einer Straftat zu beschuldigen („Sorge um Falschbeschuldigung“), eher einer Wahrheitsverzerrung unterliegen und ob Personen, die befürchten, einen Täter nicht zu entdecken („Sorge um Nicht-Aufdecken von Fehlverhalten“), eher einer Lügenverzerrung unterliegen. Nachfolgend soll die zweite Studie ausführlich beschrieben, da sie für die vorliegende Arbeit von besonderem Interesse ist.

Der Versuchsablauf dieser Studie war in vier Schritte unterteilt:

(1) Zunächst hatten die Probanden die Aufgabe, vier auf Video aufgenommene Verdächtige hinsichtlich ihrer Glaubwürdigkeit zu beurteilen. Bei den Aufnahmen handelte es sich um Aussagen von Hundebesitzern, deren Hunde verdächtigt wurden, eine Person angegriffen zu haben. Erfasst werden sollte durch diese Aufgabe eine *natürliche Urteilsverzerrung*, d.h. eine

individuelle Urteilsverzerrung der Probanden, die ohne experimentelle Variation vorhanden war. Dabei waren die Videoaufnahmen jeweils in vier Sequenzen geteilt. Die Probanden sollten nach der ersten Sequenz jeweils auf einer Skala von 1 (*Person erzählt definitiv die Wahrheit*) bis 6 (*Person lügt definitiv*) ein erstes Glaubwürdigkeitsurteil abgeben. Nach der Beurteilung der ersten Sequenz konnten sie sich zudem entscheiden, ob sie noch weitere Sequenzen ansehen wollten. Dabei wurde den Probanden mitgeteilt, dass sie mit jeder weiteren Sequenz, die sie anschauten, weniger Punkte für ein akkurates, finales Urteil erhielten. Anschließend sollten sie ein finales, binäres Glaubwürdigkeitsurteil (wahr vs. erfunden) über den Verdächtigen abgeben. Die *natürliche Urteilsverzerrung* wurde gemessen, indem die vier Glaubwürdigkeitsurteile der jeweils ersten Sequenzen gemittelt wurden. Des Weiteren konnte anhand dieser Aufgabe die *natürliche Fehlergewichtung* gemessen werden, d.h. die Gewichtung von Falschbeschuldigungen und Nicht-Aufdecken von Fehlverhalten, die ohne experimentelle Variation vorhanden war. Die Fehlergewichtung wurde erfasst, indem die Differenz zwischen der durchschnittlichen Anzahl an gesehenen Videosequenzen vor einem „wahr“-Urteil und der durchschnittlichen Anzahl an gesehenen Videosequenzen vor einem „erfunden“-Urteil berechnet wurde.

(2) Sodann erfolgte die *Manipulation der Urteilsverzerrung*. Dazu wurde den Probanden mitgeteilt, dass es in der Vergangenheit viele falsche Verurteilungen bei Angriffen von Hunden gegeben hatte. In der Bedingung „Sorge um Falschbeschuldigung“ wurde den Probanden mitgeteilt, dass mehrfach Hundebesitzer, deren Hund tatsächlich niemanden angefallen hatte,

unschuldig verurteilt worden waren; in der Bedingung „Sorge um Nicht-Aufdecken von Fehlverhalten“ wurden sie hingegen informiert, dass Hundebesitzer, deren Hund tatsächlich jemanden angefallen hatte, ungestraft davon gekommen waren. Des Weiteren wurde den Probanden in Abhängigkeit der Bedingung physische, finanzielle und mentale Konsequenzen entweder von unschuldig verurteilten Hundebesitzern oder attackierten Opfern aufgelistet.

(3) Anschließend wurde die *zentrale abhängige Variable* erfasst. Hierbei sollten die Probanden sich zwei – diesmal auf Audiokassette aufgenommene – Aussagen anhören, auf denen Hundebesitzer zu hören waren, die eine Attacke ihres Hundes abstritten. Die Probanden wurden dann gebeten, die *Schuld des Hundebesitzers* einzustufen und anzugeben, ob dieser die Wahrheit sagt oder lügt.

(4) Zum Schluss erfolgte ein Manipulationscheck. Dabei wurden die Probanden gebeten, die Schwere der beiden Fehlerarten einzustufen.

Wie erwartet deuten die Ergebnisse darauf hin, dass Probanden, die sich stärker um Falschbeschuldigungen sorgten und Informationen über unschuldig verurteilte Hundebesitzer erhielten (Bedingung „Sorge um Falschbeschuldigung“), eine weniger ausgeprägte Lügenverzerrung zeigten im Vergleich zu Personen, die sich mehr darum sorgten, dass Täter ungestraft davon kommen („Sorge um Nicht-Aufdecken von Fehlverhalten“). Entgegen der Hypothesen zeigten indes Probanden, die sich mehr darum sorgten, jemanden falsch zu beschuldigen, eine stärkere Lügenverzerrung, nachdem sie Informationen über ungestrafte Täter erhalten hatten. Die Autorinnen untersuchten zwar den Prozess, der Urteilsverzerrungen unterliegt, jedoch

nicht, wie sich unterschiedliche Verarbeitungsmodi in Abhängigkeit dieser Urteilsverzerrungen auf den Prozess der Urteilsbildung auswirken. Des Weiteren wurde auch nicht die Akkuratheit der Probanden beim Erkennen von Lüge und Wahrheit untersucht.

Einfluss des Zeitpunkts des Urteils auf die Urteilsverzerrung. Masip et al. (2009) untersuchten, inwiefern sich die Urteilsverzerrung und die Akkuratheit von Glaubwürdigkeitsurteilen im Laufe einer Videosequenz verändern. Ihren Probanden wurde ein Interview vorgelegt, das in mehrere Sequenzen unterteilt war. Nach jeder Sequenz wurden die Probanden gebeten, einzuschätzen, ob die Aussagen der jeweiligen Sequenz wahr oder erfunden sind.

Die Autoren gehen davon aus, dass die Probanden mit steigender Anzahl an gesehenen Videosequenzen von einer heuristischen Verarbeitung zu einer systematischen Verarbeitung wechseln. Dieser Wechsel des Verarbeitungsmodus wird dadurch erklärt, dass durch die größere Anzahl an gesehenen Videosequenzen mehr Informationen zur Verfügung stehen, auf deren Grundlage eine systematische Verarbeitung erfolgen kann. Somit nehmen sie an, dass bei kurzen Videosequenzen aufgrund der geringen Menge an Informationen lediglich eine heuristische, nicht jedoch eine systematische Verarbeitung möglich ist. Sie erläuterten weiter, dass heuristische Glaubwürdigkeitsurteile sogenannte „Urteile der Wahrhaftigkeit“ (judgments of truthfulness, eigene Übersetzung) sind, d.h. dass die Wahrheitsverzerrung eine kognitive Heuristik ist, die bei heuristischer Verarbeitung angewandt wird. Daher gehen sie davon aus, dass mit steigender Anzahl an gesehenen Sequenzen

weniger „wahr“-Urteile gefällt werden. Dies erklären sie dadurch, (1) dass der automatische erste Eindruck sich verändert und (2) dass die größere Informationsmenge ein Urteil erlaubt, dass stärker durchdacht ist. Sie gehen weiterhin davon aus, dass es insbesondere bei tatsächlich erfundenen Aussagen zu weniger „wahr“-Urteilen kommt. Dies wird durch die akkurate Identifizierung von wahren Aussagen erklärt. Des Weiteren nehmen sie an, dass durch die sinkende Anzahl an „wahr“-Urteilen die durchschnittliche Trefferquote und insbesondere die Trefferquote für erfundene Aussagen steigt.

Wie erwartet zeigen die Ergebnisse, dass Probanden bei den ersten Sequenzen einer Wahrheitsverzerrung unterlagen, d.h. dass sie häufiger das Urteil „wahr“ abgaben als das Urteil „erfunden“. Mit Betrachten weiterer Sequenzen wurde diese Wahrheitsverzerrung wie erwartet schwächer und die Probanden konnten besser erkennen, ob die Aussage wahr oder erfunden waren. Wie erwartet sank somit die Anzahl an „wahr“-Urteilen mit steigender Anzahl an gesehenen Videosequenzen. Ebenfalls erwartungsgemäß stieg die durchschnittliche Trefferquote für wahre und erfundene Aussagen mit steigender Anzahl an Videosequenzen an. Des Weiteren zeigte sich wie erwartet, dass Probanden erfundene Aussagen akkurater identifizierten, je mehr Videosequenzen sie gesehen hatten.

Nicht näher erläutert wurde jedoch, warum eine systematische Verarbeitung (im Vergleich zu einer heuristischen Verarbeitung) zu einer weniger ausgeprägten Wahrheitsverzerrung bei tatsächlich erfundenen Aussagen führen soll. Auch wenn indirekt gezeigt wurde, dass Probanden bei einer systematischen Verarbeitung eine höhere Trefferquote erzielten, fehlte

eine direkte Manipulation der Faktoren „Urteilsverzerrung“ und „Verarbeitungsmodus“.

Zusammenfassung. Obwohl die Untersuchung des Prozesses der verzerrten Glaubwürdigkeitsbeurteilung bedeutsam ist, gibt es nur wenige Untersuchungen, die sich diesem Thema bisher angenommen haben. Die Arbeit von Hubbell und Kollegen (2001) zeigt, dass die Wahrheitsverzerrung am stärksten ausgeprägt ist, wenn die Aussagen der Stimulusperson erfundenen sind, die Probanden hoch involviert sind und ihnen – nachdem sie die Videoaufnahmen gesehen hatten – mitgeteilt wurde, dass der Stimulusperson nicht zu trauen sei. Variiert wurde in der Studie von Hubbell und Kollegen (2001) neben der Urteilsverzerrung auch der Verarbeitungsmodus. Die Befunde sind jedoch nicht eindeutig und können auch keine Hinweise bezüglich des Einflusses der Urteilsverzerrung und des Verarbeitungsmodus auf die Akkuratheit des Urteils geben. Hurst-Wagner und Oswald (2011) untersuchten den Einfluss von Urteilsverzerrung auf das Urteil, wobei die Urteilsverzerrung anhand der Art der Fehlergewichtung manipuliert wurde. Die Ergebnisse zeigen, dass Personen, die um eine Falschbeschuldigung besorgt waren, eine weniger ausgeprägte Lügenverzerrung aufwiesen, wobei dieser Effekt durch eine Gewichtung des Fehlers moderiert wurde. Somit untersuchten sie, welcher Prozess der Urteilsverzerrung zugrundeliegt, ließen dabei aber den Verarbeitungsmodus und die Akkuratheit des Urteils außer Acht. Masip et al. (2009) zeigten, dass eine Wahrheitsverzerrung mit steigender Betrachtungsdauer von Videoaufnahmen abnimmt und Probanden von einer heuristischen zu einer systematischen Verarbeitung wechseln. Die Autoren

argumentieren, dass eine systematische Verarbeitung erst bei längerer Betrachtungsdauer stattfinden kann, da erst dann ausreichend Informationen für eine systematische Verarbeitung zur Verfügung stehen. Die Ergebnisse von Masip und Kollegen zeigten weiter, dass mit steigender Betrachtungsdauer auch die Trefferquote steigt, d.h. die Urteile der Probanden akkurater werden. Dies wird durch die systematische Verarbeitung der Aussagen begründet. Eine direkte Manipulation der Urteilsverzerrung und des Verarbeitungsmodus erfolgte in dieser Studie jedoch nicht.

Die vorliegende Arbeit schließt an diese Untersuchungen an und soll zum einen zeigen, wie der Prozess der verzerrten Glaubwürdigkeitsbeurteilung in Abhängigkeit des Verarbeitungsmodus stattfindet (Experiment 1), und zum anderen, wie sich die Faktoren „Urteilsziel“ und „Verarbeitungsmodus“ auf die Urteilsverzerrung und die Akkuratheit der Glaubwürdigkeitsurteile auswirken (Experiment 2 und 3).

1.7 Modell der verzerrten Informationsverarbeitung

Wie in den vorangegangenen Kapiteln erläutert wurde, kann soziale Informationsverarbeitung und somit die Glaubwürdigkeitsbeurteilung durch verschiedene Faktoren beeinflusst werden. Dabei kann zwischen Faktoren unterschieden werden, die die soziale Informationsverarbeitung verzerren, und solchen, die sich auf den Modus der Informationsverarbeitung auswirken und dadurch das Urteil beeinflussen. In Anlehnung an das HSM (Chen & Chaiken, 1999), die Befunde von Kunda (1990), Reinhard und Sporer (2008) sowie Masip et al. (2009) soll nun ein Modell der verzerrten sozialen Informationsverarbeitung in Bezug auf Glaubwürdigkeitsurteile vorgeschlagen

werden, welches Grundlage der nachfolgenden Untersuchungen ist (siehe Abbildung 1).

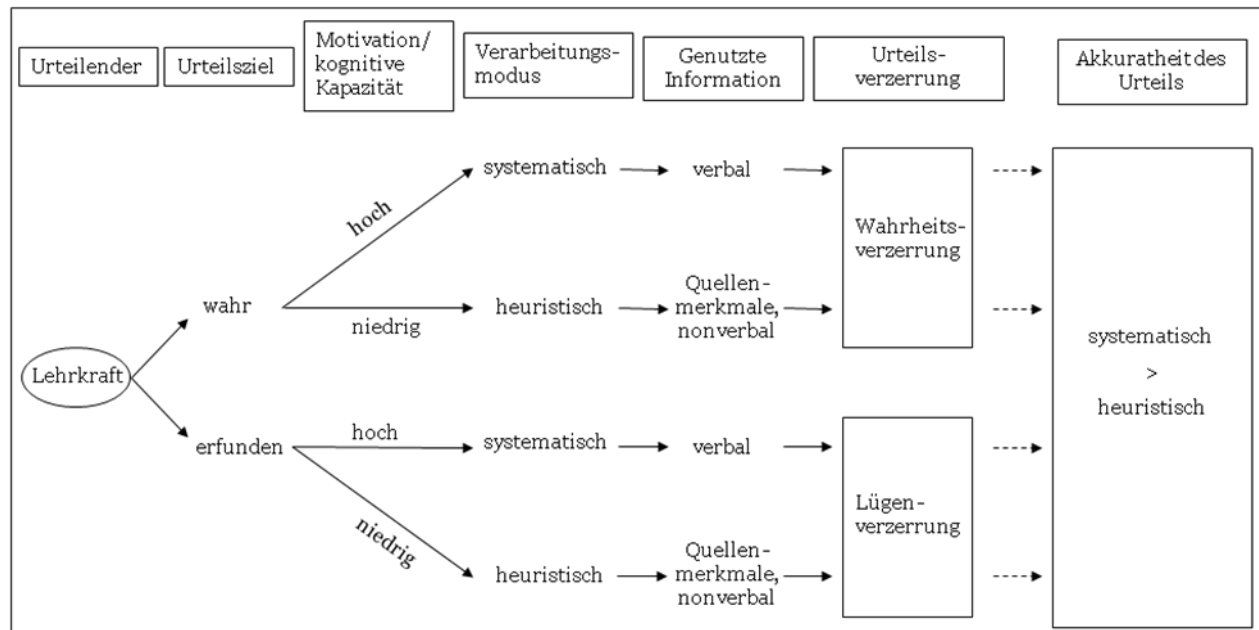


Abbildung 1. In Anlehnung an Chen und Chaiken (1999), Kunda (1990), Reinhard und Sporer (2008) und Masip et al. (2009) vorgeschlagenes Prozessmodell der verzerrten Informationsverarbeitung bei der Glaubwürdigkeitsbeurteilung.

Urteilsziel. Wie bereits oben dargestellt, wird im Modell der sozialen Informationsverarbeitung von Chen und Chaiken (1999) neben der Akkuratheitsmotivation von zwei weiteren Arten der Motivation, der Eindrucks motivation und der Verteidigungsmotivation, gesprochen. Dabei gelten die Eindrucks motivation und die Verteidigungsmotivation als Motivationen, welche die soziale Informationsverarbeitung verzerren. Kunda (1990) unterscheidet ebenfalls zwischen Faktoren, die zu einer (weitestgehend) unverzerrten und zu einer verzerrten Informationsverarbeitung führen: Den Akkuratheitszielen und den Richtungszielen. Im vorgeschlagenen Modell (Abbildung 1) werden die von Chen und Chaiken sowie Kunda vorgeschlagenen Faktoren als „Urteilsziel“ bezeichnet. *Urteilsziel* kann definiert werden als die Motivation eines Urteilers, zu einem *bestimmten* Urteil über eine andere Person

zu gelangen. Der Begriff *Urteilsziel* wurde gewählt, da er sowohl die Vorstellung von Chen und Chaiken, dass *Motivationen* die soziale Informationsverarbeitung beeinflussen, als auch die Vorstellungen von Kunda, dass *Ziele* die Verarbeitung verzerren, umfasst. Dabei wird davon ausgegangen, dass Urteilsziele auch unbewusst sein können (vgl. Uleman, 1996, 1999). Des Weiteren wird zwischen zwei Ausprägungen von Urteilszielen unterschieden: Bei einem „Urteilsziel wahr“ wird angenommen, dass Probanden motiviert sind, die Aussagen einer zu beurteilenden Person als „wahr“ zu bewerten, und somit häufiger das Urteil „wahr“ abgeben und folglich einer Wahrheitsverzerrung unterliegen. Bei einem „Urteilsziel erfunden“ wird angenommen, dass Probanden motiviert sind, die Aussagen der zu beurteilenden Person als „erfunden“ zu bewerten, und somit häufiger das Urteil „erfunden“ abgeben und damit einer Lügenverzerrung unterliegen.

Verarbeitungsmodus. Soziale Informationsverarbeitung kann wie bereits oben dargestellt nicht nur durch Urteilsziele, sondern auch durch den Modus, in dem Informationen verarbeitet werden, beeinflusst werden. In Anlehnung an bisherige Befunde und die Annahmen des HSM wird im vorliegenden Modell (Abbildung 1) zwischen dem *heuristischen Modus* und dem *systematischen Modus* der Informationsverarbeitung unterschieden. Wie bereits oben erläutert wird angenommen, dass im *heuristischen Modus* (wie auch bei Chen & Chaiken, 1999) die Aktivierung und Anwendung von Heuristiken bei der Urteilsbildung stattfindet. Es wird davon ausgegangen, dass im heuristischen Modus (in Anlehnung an die Befunde von Reinhard & Sporer, 2008, 2010) insbesondere non-verbale Informationen bzw. Quellenmerkmale (z.B. Arm- und

Handbewegungen des Senders) zur Urteilsbildung genutzt werden. Des Weiteren wird angenommen, dass im *systematischen* Modus (in Anlehnung an Chen & Chaiken, 1999) Informationen systematisch und analytisch verarbeitet werden und (in Anlehnung an die Befunde von Reinhard & Sporer, 2008) insbesondere verbale Informationen bzw. der Nachrichteninhalt (z.B. logische Konsistenz und Plausibilität der Aussagen) zur Urteilsbildung genutzt werden.

Modell der verzerrten Informationsverarbeitung bei der Glaubwürdigkeitsbeurteilung. In der vorliegenden Dissertation wird der Frage nachgegangen, wie sich – in Bezug auf Glaubwürdigkeitsurteile – unbewusste Urteilsziele und der Verarbeitungsmodus auf die Urteilsverzerrung, den Prozess der Urteilsbildung und die Akkuratheit dieser Urteile auswirken. Dabei gelten die Urteilsverzerrung und die Akkuratheit dieser Urteile als zentrale abhängige Variablen und sowohl Urteilsziel als auch Verarbeitungsmodus als Einflussgrößen. Auf Grundlage der vorgestellten Befunde wird ein Modell vorgeschlagen, dass unter Berücksichtigung der beiden Einflussvariablen *Urteilsziel* und *Verarbeitungsmodus* die (verzerrte) Verarbeitung von Informationen bei der Glaubwürdigkeitsbeurteilung darstellt (Abbildung 1). Es wird angenommen, dass bei einem „Urteilsziel wahr“ Personen – unabhängig vom Verarbeitungsmodus – eine Wahrheitsverzerrung zeigen und dass bei einem „Urteilsziel erfunden“ Personen – ebenfalls unabhängig vom Verarbeitungsmodus – dazu tendieren, Aussagen als erfunden einzustufen (vgl. Chen & Chaiken, 1999, und Kunda, 1990). Ist die Motivation hoch, wird davon ausgegangen, dass Informationen systematisch verarbeitet werden (Chen & Chaiken, 1999) und insbesondere verbale Informationen bzw. der

Nachrichteninhalt das Urteil beeinflussen (Reinhard & Sporer, 2008). Ist die Motivation dagegen niedrig, wird angenommen, dass Informationen heuristisch verarbeitet werden (Chen & Chaiken, 1999) und insbesondere nonverbale Informationen bzw. Quellenmerkmale zur Urteilsbildung genutzt werden (Reinhard & Sporer, 2008). Es wird vorhergesagt, dass unabhängig vom Urteilsziel die Akkuratheit des Glaubwürdigkeitsurteils bei systematischer Verarbeitung höher ist als bei heuristischer Verarbeitung (vgl. Masip et al., 2009). Dabei wird angenommen, dass bei einer systematischen Verarbeitung die Akkuratheit höher ist, da insgesamt mehr verbale (vs. nonverbale) Hinweisreize genutzt werden, die ihrerseits als valide Lügenindikatoren angesehen werden (DePaulo et al., 2003; Hartwig & Bond, 2011). Des Weiteren wird davon ausgegangen, dass es unabhängig vom Urteilsziel und dem Verarbeitungsmodus zu einer Wahrheitsverzerrung kommt, wenn Personen⁷ bereits einer „natürlichen“ Wahrheitsverzerrung unterliegen (vgl. Hubbell et al., 2001; Reinhard, Dickhäuser, et al., 2011). Dies sollte auch dazu führen, dass die Akkuratheit des Urteils unabhängig vom Urteilsziel und Verarbeitungsmodus höher bei tatsächlich wahren im Vergleich zu erfundenen Aussagen ist (Wahrheitseffekt; Levine et al., 1999). Weiterhin wird angenommen, dass bei einem „Urteilsziel wahr“ unabhängig vom Verarbeitungsmodus mehr wahre Aussagen akkurat erkannt werden und bei einem „Urteilsziel erfunden“ mehr erfundene Aussagen richtig eingeschätzt werden (vgl. C. F. Bond & DePaulo, 2006). Diese aus dem Modell abgeleiteten Annahmen sollen in der vorliegenden Arbeit überprüft werden.

⁷ In der vorliegenden Arbeit handelt es sich um (angehende) Lehrkräfte.

2 Pilotstudie 1: Der Einfluss von Urteilszielen mit angenommenen Urteilskonsequenzen auf Glaubwürdigkeitsurteile

In der vorliegenden Pilotstudie 1 wurde in Anlehnung an Hurst-Wagner und Oswald (2011) eine experimentelle Manipulation des Urteilsziels getestet, die insbesondere für den pädagogischen Bereich von Bedeutung ist. So wurde untersucht, inwiefern die Motivation, bestimmte (Urteils-)Fehler zu vermeiden (Urteilsvermeidung), sich auf den Prozess der Glaubwürdigkeitsbeurteilung auswirkt und das Glaubwürdigkeitsurteil verzerrt (Urteilsverzerrung). Wie bereits oben dargestellt wurde, unterliegen Personen, die besorgt sind, jemanden fälschlicherweise einer Straftat zu beschuldigen, eher einer Wahrheitsverzerrung und Personen, die besorgt sind, einen Täter nicht zu entdecken, eher einer Lügenverzerrung (Hurst-Wagner & Oswald, 2011). Aufgrund der Ausbildung einer Lehrkraft (vgl. Terhart, 2005, Standards für die Lehrerbildung, insb. Kompetenzbereich „Erziehen“) wird angenommen, dass Lehrkräfte zum einen motiviert sind, Fehlverhalten von Schülern (z.B. das Nutzen von unerlaubten Hilfsmitteln bei einer Klassenarbeit) aufzudecken, und zum anderen voreilige Anschuldigungen bezüglich eines Fehlverhaltens zu vermeiden (z.B. um eine positive Lehrer-Schüler-Interaktion nicht zu gefährden). Es wird angenommen, dass aufgrund der Motivation, *Fehlverhalten von Schülern aufzudecken*, Lehrkräfte dazu tendieren, einem Täuschungsverdacht eher nachzugehen, auch wenn dies die Tendenz nach sich zieht, dass den Aussagen von Schülern von der Lehrkraft eher misstraut wird. Dieses Misstrauen wiederum sollte dann zu einer so genannten Lügenverzerrung führen, d.h. zu der Tendenz der Lehrkraft, Aussagen von

Schülern eher als erfunden einzustufen (vgl. Kapitel 1.4.3 und Vrij, 2008).

Weiterhin wird angenommen, dass die Motivation, *Falschbeschuldigungen zu vermeiden*, dazu führt, dass Lehrkräfte dazu tendieren, einem Täuschungsverdacht nicht nachzugehen; es wird vorhergesagt, dass dies die Tendenz nach sich zieht, dass die Lehrkraft den Aussagen von Schülern eher vertraut. Vertrauen wiederum soll dann zu einer Wahrheitsverzerrung führen, d.h. zu der Tendenz der Lehrkraft, Aussagen eher als „wahr“ einzustufen (vgl. Kapitel 1.4.3 und Vrij, 2008).

2.1 Fragestellung und Hypothesen

In Pilotstudie 1 wurde untersucht, wie sich Urteilsziele auf den Prozess der Glaubwürdigkeitsbeurteilung und die Urteilsverzerrung auswirken.

Urteilsziele wurden durch die Motivation, bestimmte Fehler bzw. Urteile zu vermeiden (Urteilsvermeidung: wahr vs. erfunden), experimentell variiert. Es wurde angenommen, dass Probanden das Urteil „erfunden“ vermeiden und somit eher das Urteil „wahr“ abgeben, wenn sie motiviert sind,

Falschbeschuldigungen zu vermeiden (Bedingung „Vermeidung Urteil erfunden“). Dagegen sollten Probanden das Urteil „wahr“ vermeiden und somit eher das Urteil „erfunden“ abgeben, wenn sie motiviert sind, Fehlverhalten aufzudecken (Bedingung „Vermeidung Urteil wahr“). Weiterhin wurde angenommen, dass Probanden der Bedingung „Vermeidung Urteil erfunden“ einer Wahrheitsverzerrung unterliegen und Probanden der Bedingung „Vermeidung Urteil wahr“ einer Lügenverzerrung unterliegen.

2.2 Methode

2.2.1 Versuchsaufbau

Die Daten der Pilotstudie 1 wurden mit Hilfe der Erhebungssoftware Unipark (www.unipark.info) erhoben. Die Studierenden wurden zunächst begrüßt, über die zu erwartende Aufwandsentschädigung für die Teilnahme informiert und kurz über den Untersuchungsablauf aufgeklärt. Anschließend wurden demographische Variablen (Alter, Geschlecht, Studiengang, Fachsemesterzahl) erfasst. Aufgrund der Befunde der bereits oben dargestellten Studie (Kapitel 1.3.2) von Marksteiner, Reinhard, et al. (2013) wurde als Studienkontext das Mobbing gewählt. Hierzu wurden die Probanden gebeten, sich in eine Situation hineinzusetzen, in der sie Lehrer an einer neuen Schule sind und in welcher der Schüler eines Kollegen einen Brief mit beleidigendem Inhalt in seiner Tasche findet. Der Schüler fühlt sich aufgrund des Briefes, der an ihn adressiert war, bedroht und äußert die Vermutung, dass ein Mitschüler den Brief in seine Tasche steckte, als diese während einer Pause unbeaufsichtigt war. Danach folgte ein kurzer Ausschnitt des Briefes (siehe Anhang I.a). Dieser Ausschnitt stammt aus einer Studie von Dickhäuser et al. (2012). Weiterhin sollten sich die Probanden vorstellen, dass ihr Kollege sich verpflichtet fühlt herauszufinden, wer an der Verfassung des Briefes beteiligt war, und plant, verschiedene Mitschüler, die sich während der besagten Pause in der Nähe der Tasche aufhielten, zu befragen. An dieser Stelle werden die Probanden gebeten sich vorzustellen, dass sie bei einem der Gespräche zwischen ihrem Kollegen und einem der Mitschüler anwesend sind. Danach folgte die Manipulation des Urteilsziels und ein Manipulationscheck, bei dem die Probanden gebeten

wurden auf einer Skala von 1 bis 7 anzugeben, ob sie es für schlimmer halten, Fehlverhalten von Schülern bei Mobbing nicht aufzudecken oder Schüler zu Unrecht eines Fehlverhaltens bei Mobbing zu verdächtigen. Die beiden Endpunkte der Skala waren „Es ist schlimmer, Fehlverhalten von SchülerInnen bei Mobbing nicht aufzudecken.“ (1) und „Es ist schlimmer, SchülerInnen zu Unrecht zu verdächtigen, an einem Mobbing-Vorfall beteiligt gewesen zu sein.“ (7). Anschließend wurde den Probanden das Gespräch zwischen dem besagten Kollegen und einem der Mitschüler vorgelegt. Danach wurde das Glaubwürdigkeitsurteil über die Aussagen des Schülers auf einer Skala von 1 (*erlogen*) bis 10 (*wahr*) erfasst. Am Ende wurde den Probanden für ihre Teilnahme gedankt und über das tatsächliche Ziel der Untersuchung aufgeklärt.

2.2.2 Versuchsmaterial und Vortest des Gesprächs

Manipulation des Urteilsziels durch Urteilsvermeidung. Die Manipulation des Urteilsziels erfolgte mithilfe des *Priming-Effekts* (Higgins, Rholes, & Jones, 1977). Als *Priming* „bezeichnet man den Prozess, bei dem vorhergehende Erfahrungen die Zugänglichkeit eines Schemas, einer Persönlichkeitseigenschaft oder eines Konzepts erhöhen“ (Aronson et al., 2008). Probanden in der Bedingung „Vermeidung Urteil erfunden“ wurden Konsequenzen zugänglich gemacht, die durch Falschbeschuldigung entstehen können. Dagegen wurden Probanden der Bedingung „Vermeidung Urteil wahr“ Konsequenzen aufgezeigt, die durch Nicht-Aufdecken von Fehlverhalten entstehen können. Der genaue Wortlaut für die Bedingung „Vermeidung Urteil erfunden“ war wie folgt:

„Bitte beachten Sie, dass es schwerwiegende Konsequenzen haben kann, SchülerInnen zu Unrecht zu verdächtigen, an einem Mobbing-Vorfall beteiligt gewesen zu sein. Falschbeschuldigung bei Mobbing kann folgende schwerwiegende Konsequenzen haben:

- Zum einen für das Selbstkonzept der SchülerInnen, da SchülerInnen, die fälschlicherweise beschuldigt werden, einen Mobbing-Brief verfasst zu haben, annehmen, dass ihnen ein derartiges Verhalten zugetraut wird. Dies hat negative Konsequenzen auf das Selbstkonzept des Schülers.*
- Zum anderen erhöht eine Falschbeschuldigung die Wahrscheinlichkeit, dass SchülerInnen ihr Verhalten in eine nicht erwünschte, negative Richtung ändern.*
- Weiterhin können Falschbeschuldigungen sich negativ auf das Schüler-Lehrer-Vertrauensverhältnis auswirken, da Lehrkräfte von SchülerInnen, die fälschlicherweise des Verfassens eines Mobbing-Briefes beschuldigt werden, als ungerecht und unfair bewertet werden.“*

Der genaue Wortlaut für die Bedingung „Vermeidung Urteil wahr“ war ähnlich aufgebaut wie der Wortlaut in der Bedingung „Vermeidung Urteil erfunden“ und lautete wie folgt:

„Bitte beachten Sie, dass es schwerwiegende Konsequenzen haben kann, wenn bei SchülerInnen Täuschungen bei Mobbingfällen nicht erkannt wird. Das Nicht-Aufdecken von Fehlverhalten bei Mobbing kann folgende schwerwiegende Konsequenzen haben:

- *Zum einen für Mobbing-Opfer, da Mobbing durch Peers Auswirkungen auf psychologische Variablen haben kann und zu Depressionen und in seltenen Fällen zu Selbstmord führen kann.*
- *Zum anderen kommen beim Nicht-Aufdecken von Fehlverhalten bei Mobbing die schuldigen SchülerInnen ungestraft ‚davon‘. Dies erhöht die Wahrscheinlichkeit, dass SchülerInnen ihr negatives Verhalten fortsetzen und verstärken.*
- *Weiterhin kann das Nicht-Aufdecken von Fehlverhalten bei Mobbing sich negativ auf das Schüler-Lehrer-Verhältnis auswirken, da Lehrkräfte von SchülerInnen, deren Fehlverhalten in Bezug auf Mobbing unentdeckt bleibt, als unaufmerksam und ‚laissez-faire‘ bewertet werden.“*

Beschreibung des Gesprächs. Als Information diente ein Gespräch zwischen einem Schüler (sog. „Schüler S.“) und dem besagten Kollegen (siehe Anhang I.b). Als Grundlage für die Entwicklung des Gesprächs diente die Transkription der Videoaufnahme eines Schülers aus der Studie von Dickhäuser et al. (2012). Aufgrund der Ergebnisse dieser Studie zu Vorstellungen über Begleiterscheinungen von Lügen bei angehenden Lehrkräften wurde das transkribierte Gespräch leicht verändert und verschiedene subjektive Lügenindikatoren eingebaut (vgl. Bond, 2006). Subjektive Lügenindikatoren können verstanden werden als Indikatoren, von denen *angenommen* wird, dass sie Hinweise darauf liefern, ob eine Person die Wahrheit sagt oder lügt. Die in das Gespräch eingebauten Lügenindikatoren waren „logische Inkonsistenz“, „Wortwiederholungen“, „Details“ und „Erwähnen von Verantwortung“. Die

Befunde von Marksteiner, Reinhard, et al. (2012) deuten darauf hin, dass Urteilende auf diese Hinweise achten, um ein Glaubwürdigkeitsurteil abzugeben. Die ersten beiden Hinweise sollten dazu führen, dass die Aussagen des Schülers als eher unglaubwürdig wahrgenommen wurden. So widerspricht sich der Schüler während des Gesprächs (Hinweis „logische Inkonsistenz“) und wiederholt das Wort „Kinofilme“ insgesamt dreimal (Hinweis „Wortwiederholungen“). Die letzten beiden Hinweise sollten die Glaubwürdigkeit der Aussagen des Schülers erhöhen. So beinhalteten die Aussagen des Schülers Details (Hinweis „Details“) und der Schüler lehnt es ab, Verantwortung für den Vorfall zu übernehmen (Hinweis „Erwähnen von Verantwortung“).

Vortest des Gesprächs. In einem Vortest wurde untersucht, ob das vorliegende Versuchsmaterial als mehrdeutig bzw. ambivalent eingestuft werden kann.⁸ Es wurde angenommen, dass das Gespräch als durchschnittlich glaubwürdig eingestuft wurde. Weiterhin wurde angenommen, dass die Probanden des Vortests u.a. die eingebauten Hinweisreize als Argumente für ihr Glaubwürdigkeitsurteil heranziehen würden. An dem Vortest nahmen $N = 21$ Studierende der Universität Mannheim teil (66.7% weiblich, 52.4% Lehramtsstudierende, $M_{\text{Alter}} = 23.76$ Jahre, $SD_{\text{Alter}} = 2.95$). Die Studierenden wurden gebeten, das Gespräch zwischen Schüler S. und seinem Lehrer über den Mobbing-Brief zu lesen und anschließend die Glaubwürdigkeit des Schülers auf einer 10-Punkte-Likert-Skala (1 = unglaubwürdig, 10 = glaubwürdig)

⁸ Das Versuchsmaterial sollte ambivalent und nicht eindeutig sein, damit die Probanden mehr Spielraum zur Interpretation des Gesprächs hatten (vgl. Giner-Sorolila & Chaiken, 1999).

einzustufen. Im Anschluss daran wurden sie gebeten, zum einen Argumente zu nennen, die für die Glaubwürdigkeit des Schülers sprechen (Pro-Schüler-Argumente), und zum anderen Argumente, die gegen die Glaubwürdigkeit sprechen (Contra-Schüler-Argumente). Das mittlere Glaubwürdigkeitsurteil betrug $M = 6.10$ ($SD = 2.36$) und unterschied sich nicht signifikant vom Skalenmittelpunkt 5.5, $t(20) = 1.15$, $p = .26$. Somit kann das Gespräch als nicht eindeutig einstuftbar bzw. als ambivalent angesehen werden. Auf Ambivalenz deutet auch die relativ hohe Standardabweichung hin. Wie erwartet wurden neben anderen Argumenten die vier Hinweise „logische Inkonsistenz“, „Wortwiederholungen“, „Details“ und „Erwähnen von Verantwortung“ als Argumente für das Glaubwürdigkeitsurteil angeführt.

2.2.3 *Versuchsdesign*

Pilotstudie 1 folgte einem einfaktoriellen Design mit dem Faktor Urteilsvermeidung und den Faktorstufen „Vermeidung Urteil wahr“ und „Vermeidung Urteil erfunden“.

2.2.4 *Stichprobe*

An der Untersuchung nahmen 36 Studierende (77.8% weiblich) teil. Alle Studierenden waren angehende Lehrkräfte. Das Durchschnittsalter lag bei 22.97 Jahren ($SD = 3.70$), die Probanden befanden sich durchschnittlich im 5. Fachsemester ($SD = 2.70$).

2.3 Ergebnisse

2.3.1 Manipulationscheck

Zunächst wurde getestet, ob sich die Manipulation des Urteilsziels darauf auswirkt, ob die Vermeidung einer Falschbeschuldigung oder das Nicht-Aufdecken von Fehlverhalten als schlimmer eingestuft wird. Ein Kolmogorov-Smirnov-Test auf Normalverteilung zeigt, dass die Verteilung der abhängigen Variable „Fehlervermeidung“ nicht in beiden Experimentalgruppen normalverteilt ist (Vermeidung Urteil erfunden: $p = .20$; Vermeidung Urteil wahr: $p < .001$). Da Varianzanalysen bzw. F -Statistiken jedoch sehr robust gegenüber Verletzungen ihrer Annahmen sind (vgl. Field, 2009), wurde dennoch eine ANOVA berechnet.⁹ Eine einfaktorielle ANOVA mit der abhängigen Variable „Fehlervermeidung“ und der unabhängigen Variable „Urteilsvermeidung“ zeigte einen signifikanten Effekt des Faktors Urteilsvermeidung, $F(1, 34) = 17.20, p < .001$. Wie erwartet stuften Probanden der Bedingung „Vermeidung Urteil erfunden“ es als schlimmer ein, einen Schüler des Mobbing fälschlicherweise zu beschuldigen ($M = 4.21, SD = 1.75$) als Probanden der Bedingung „Vermeidung Urteil wahr“ ($M = 2.06, SD = 1.30$); diese tendierten dazu, das Nicht-Aufdecken von Fehlverhalten bei Mobbing als

⁹ Die Ergebnisse des Mann-Whitney-Tests deuten darauf hin, dass Probanden der Bedingung „Vermeidung Urteil erfunden“ ($Mdn = 4.00$) sich signifikant von den Probanden der Bedingung „Vermeidung Urteil wahr“ ($Mdn = 1.00$) unterschieden, $U = 54.5, z = -3.46, p < .001$, und eher dazu tendierten, dass es wichtig sei, keine Falschbeschuldigungen vorzunehmen. Die Befunde deuten ebenfalls darauf hin, dass die Manipulation des Urteilsziels erfolgreich war.

schlimmer einzustufen. Die Befunde deuten darauf hin, dass die Manipulation des Urteilsziels erfolgreich war.

2.3.2 Glaubwürdigkeitsurteil

Im Mittel stuften die Probanden die Aussagen des Schülers als $M = 6.33$ ($SD = 2.63$) glaubwürdig ein. Dieser Wert unterscheidet sich marginal signifikant vom Skalenmittelwert 5.5, $t(35) = 1.90$, $p = .065$. Dies deutet darauf hin, dass die Probanden die Aussagen von Schüler S. durchschnittlich eher als glaubwürdig einstufen. Um zu testen, ob die Urteilsvermeidung sich auf die Beurteilung der Glaubwürdigkeit der Aussagen des Schülers S. auswirkt, wurde eine ANOVA durchgeführt.¹⁰ Die Ergebnisse der einfaktoriellen ANOVA mit der abhängigen Variable „Glaubwürdigkeit“¹¹ und der unabhängigen Variable „Urteilsvermeidung“ deuten wider Erwarten darauf hin, dass die Urteilsvermeidung keinen Einfluss auf die Einstufung der Glaubwürdigkeit der Schüleraussagen hat ($F < 1$).¹²

¹⁰ Ein Kolmogorov-Smirnov-Test deutet darauf hin, dass die Variable „Glaubwürdigkeit“ in beiden Experimentalgruppen nicht normalverteilt ist (Urteilsvermeidung erfunden: $p < .001$; Urteilsvermeidung wahr: $p = .001$). Der Levene-Test auf Varianzhomogenität deutet darauf hin, dass die Varianzen der beiden Experimentalgruppen gleich sind, $F(1, 34) = 0.32$, $p = .57$.

¹¹ Strenggenommen handelt es sich hierbei um eine kategoriale Variable, genauer gesagt um eine ordinale Variable; häufig werden jedoch Daten, die auf Selbstberichten beruhen (und strenggenommen ordinal sind), wie kontinuierliche Variablen behandelt (vgl. Field, 2009).

¹² Die Ergebnisse eines Mann-Whitney-Tests deuten darauf hin, dass Probanden der Bedingung „Vermeidung Urteil erfunden“ ($Mdn = 8.00$) die Aussagen des Schülers *nicht* als signifikant glaubwürdiger einstufen als Probanden der Bedingung „Vermeidung Urteil wahr“ ($Mdn = 7.00$), $U = 140.5$, $z = -0.68$, $p = .51$.

2.4 Diskussion

Obwohl der Faktor „Urteilsvermeidung“ einen signifikanten Effekt darauf hat, welcher Fehler eher vermieden werden sollte (d.h. ob es schlimmer ist, einen Schüler bei Mobbing fälschlicherweise zu beschuldigen oder sein Fehlverhalten nicht aufzudecken), beeinflusste der Faktor „Urteilsvermeidung“ das Glaubwürdigkeitsurteil nicht. Erklärt werden kann dieser Befund dadurch, dass das Urteil keine realen Konsequenzen nach sich zog. So wurden theoretische Konsequenzen der Fehler „Falschbeschuldigung bei Mobbing“ und „Nicht-Aufdecken von Fehlverhalten bei Mobbing“ den Probanden per Priming zugänglich gemacht, allerdings gab es für die Probanden keine realen Konsequenzen in Abhängigkeit ihres Urteils. In einer realen Situation, in welcher der Lehrer einen Schüler entweder fälschlicherweise eines Fehlverhaltens beschuldigt oder ein Fehlverhalten nicht aufdeckt, können jedoch tatsächliche Konsequenzen folgen. In einer hypothetischen Situation, wie sie in einem Experiment vorliegt, in welcher der Proband keine Konsequenzen seines Urteils erfährt, scheint das durch Priming manipulierte Urteilsziel dagegen keinen Effekt auf das Glaubwürdigkeitsurteil zu haben. Daher sollte eine Manipulation des Urteilsziels gewählt werden, bei der das Urteil reale, tatsächliche Konsequenzen nach sich zieht, die den Probanden direkt treffen und keine Konsequenzen, die theoretisch und nach einem längeren Zeitraum aufgrund eines bestimmten Urteils entstehen können.

3 Pilotstudie 2: Der Einfluss von unbewussten Urteilszielen mit realen Urteilskonsequenzen auf Glaubwürdigkeitsurteile

3.1 Fragestellung und Hypothesen

Ziel der Pilotstudie 2 war es zu überprüfen, ob ein Urteilsziel mit realen Urteilskonsequenzen die Verarbeitung von Informationen und das Glaubwürdigkeitsurteil beeinflussen. Das Urteilsziel wurde mit Hilfe der Motivation manipuliert, wonach Personen bei der Bearbeitung einer Aufgabe (z.B. die Beantwortung von Fragen bei einem Experiment) ökonomisch handeln, d.h. versuchen, ihren Aufwand so minimal wie möglich zu halten (z.B. Gesetz der wenigen Arbeit [Law of Less Work], Hull, 1943; Kognitiver Geizhals [cognitive miser], Taylor, 1981). Daher kann angenommen werden, dass Personen in einem experimentellen Setting zu denjenigen Urteilen tendieren, die mit weniger Aufwand verbunden sind. So sollten Personen der Bedingung „Vermeidung Urteil erfunden“, für die es ökonomisch sinnvoller ist, zu dem Urteil zu kommen, dass die zu beurteilende Person *die Wahrheit sagt*, eher das Urteil „wahr“ abgeben als Personen der Bedingung „Vermeidung Urteil wahr“, für die es ökonomisch sinnvoller wäre, das Urteil zu fällen, dass die Aussagen der zu beurteilenden Person *erfunden* sind (Hypothese 1). Des Weiteren sollte untersucht werden, ob Urteilsziele die Verarbeitung von urteilsrelevanten Informationen in Richtung des Urteilsziels verzerren. Es wurde angenommen, dass die Informationen in verzerrter Art in Abhängigkeit vom Urteilsziel verarbeitet werden (vgl. Kunda, 1990, 1999). So sollten Personen der Bedingung „Vermeidung Urteil erfunden“ mehr Argumente generieren, die für die Unschuld der Zielperson bei einem Täuschungsvorfall sprechen. Dagegen

sollten Personen der Bedingung „Vermeidung Urteil wahr“ mehr Argumente nennen, welche für die Schuld der Zielperson sprechen (Hypothese 2).

3.2 Methode

3.2.1 Versuchsablauf

Pilotstudie 2 wurde mit einem Papierfragebogen erhoben. Zunächst erhielten die Probanden allgemeine Informationen zur Untersuchung (z.B. über Dauer und Aufwandsentschädigung) und beantworteten demographische Fragen bezüglich ihres Alters, Geschlechts, Studiengangs und Fachsemesters. Danach wurden sie gebeten, sich in die Situation des Mobbing-Vorfalles (wie in Pilotstudie 1 beschrieben) hineinzusetzen. Anschließend folgte die Manipulation des Urteilsziels. Wie in Pilotstudie 1 wurden die Probanden in beiden Bedingungen darüber informiert, dass sie ein Interview lesen würden, in dem ein Kollege einen Schüler befragt, der verdächtigt wird, an der Verfassung eines Briefes beteiligt gewesen zu sein. Des Weiteren wurden sie informiert, dass sie die Glaubwürdigkeit der Aussagen des befragten Schülers auf einer Skala von 1 (*erfunden*) bis 10 (*wahr*) einschätzen sollten und Argumente für ihr Urteil abgeben müssten. Den Probanden wurde mitgeteilt, dass die Werte von 1 bis 5 bedeuteten, dass die Aussagen als erfunden eingestuft wurden, und die Werte von 6 bis 10, dass die Aussagen als wahr eingestuft wurden. Anschließend wurde das Urteilsziel manipuliert: In der Bedingung „Vermeidung Urteil erfunden“ wurden die Probanden darüber informiert, dass sie einen zweiseitigen Fallbericht über den befragten Schüler verfassen müssten, wenn sie die Aussagen des Schülers für erfunden hielten (d.h. wenn sie auf der Glaubwürdigkeitsskala einen Wert zwischen 1 und 5 markierten). Sofern sie die

Aussagen für wahr hielten, d.h. wenn sie auf der Skala einen Wert zwischen 6 und 10 markierten, mussten sie keinen Fallbericht schreiben. In der Bedingung „Vermeidung Urteil wahr“ wurden die Probanden hingegen darüber informiert, dass sie einen zweiseitigen Fallbericht über den befragten Schüler verfassen mussten, wenn sie die Aussagen des Schülers für wahr hielten (d.h. wenn sie auf der Glaubwürdigkeitsskala einen Wert zwischen 6 und 10 markierten). Sofern sie die Aussagen dagegen für erfunden hielten (d.h. wenn sie auf der Skala einen Wert zwischen 1 und 5 markierten), mussten sie keinen Fallbericht schreiben. Die Probanden mussten folglich weniger Zeit und Aufwand aufbringen, wenn sie in der Bedingung „Vermeidung Urteil erfunden“ die Aussagen als wahr und in der Bedingung „Vermeidung Urteil wahr“ die Aussagen als erfunden einstufen. Das Schreiben des zweiseitigen Fallberichts wurde damit als reale Konsequenz des Glaubwürdigkeitsurteils angesehen. Nach der Manipulation des Urteilsziels wurde den Probanden das Interview zwischen dem besagten Kollegen und dem Schüler vorgelegt. Anschließend stuften die Probanden die Glaubwürdigkeit der Aussagen des Schülers ein und gaben Argumente für ihr Urteil an. Sie wurden gebeten, die Argumente für ihr Urteil durchzunummerieren. Danach folgte ein Instruktionscheck (siehe Oppenheimer, Meyvis, & Davidenko, 2009, für weitere Informationen über Instruktionschecks), bei dem die Probanden gebeten wurden, anzugeben, unter welchen Umständen sie einen zweiseitigen Fallbericht schreiben sollten. Die Antwortoptionen waren „... wenn ich glaube, dass die Aussagen des Schülers wahr sind.“; „... wenn ich glaube, dass die Aussagen des Schülers erfunden sind.“ und „Ich weiß es nicht.“. Zum Schluss wurde den Probanden für ihre

Teilnahme gedankt, sie wurden gebeten – sofern erforderlich –, den zweiseitigen Fallbericht zu verfassen und über den tatsächlichen Inhalt der Studie aufgeklärt.

3.2.2 *Versuchsmaterial*

Das Versuchsmaterial war identisch zum Versuchsmaterial in Pilotstudie 1.

3.2.3 *Versuchsdesign*

Pilotstudie 2 folgte einem einfaktoriellen Design mit dem Faktor „Urteilsvermeidung“ und den Faktorstufen „Vermeidung Urteil wahr“ und „Vermeidung Urteil erfunden“.

3.2.4 *Stichprobe*

An Pilotstudie 2 nahmen 46 angehende Lehrer teil. Neun Teilnehmer wurden von weiteren Analysen ausgeschlossen, da sie entweder den Instruktionscheck nicht bestanden ($N = 4$), den zweiseitigen Fallbericht nicht verfassten, obwohl sie ihn aufgrund ihres Urteils hätten schreiben sollen ($N = 4$), oder weder den Instruktionscheck bestanden noch den zweiseitigen Fallbericht – obwohl erfordert – nicht verfassten ($N = 1$). Mehr als die Hälfte der Probanden war weiblich (75.5%), was für Lehramtsstudierende als gewöhnliche Verteilung angesehen werden kann. Im Durchschnitt waren die Probanden 22.7 ($SD = 2.49$) Jahre alt und studierten im 5.30 ($SD = 2.68$) Semester.

3.3 Ergebnisse

3.3.1 *Glaubwürdigkeitsurteil*

Im Durchschnitt wurden die Aussagen des Schülers als $M = 5.05$ ($SD = 2.19$) glaubwürdig eingestuft. Dieser Wert unterschied sich nicht signifikant vom Skalenmittelwert 5.5, $p = .22$. Eine ANOVA¹³ mit dem Glaubwürdigkeitsurteil als abhängiger Variable und der Urteilsvermeidung als unabhängiger Variable zeigte einen signifikanten Haupteffekt der Urteilsvermeidung, $F(1, 35) = 8.93$, $p = .005$, $\eta^2 = 0.203$. Somit beurteilten die Probanden in der Bedingung „Vermeidung Urteil erfunden“ die Aussagen als signifikant glaubwürdiger ($M = 6.00$, $SD = 1.92$) als die Probanden der Bedingung „Vermeidung Urteil wahr“ ($M = 4.06$, $SD = 2.04$).

3.3.2 *Binäres Glaubwürdigkeitsurteil*

Um zu überprüfen, ob Probanden in Abhängigkeit ihres Urteilsziels dazu tendieren ein bestimmtes Urteil abzugeben, wurde das kontinuierliche Glaubwürdigkeitsurteil in ein binäres Glaubwürdigkeitsurteil transformiert. Das binäre Urteil wurde berechnet, indem Probanden, die einen Wert zwischen 1 und 5 auf der 10-Punkte-Skala markierten, der Wert 1 zugewiesen wurde, und Probanden, die einen Wert zwischen 6 und 10 markierten, den Wert 2 erhielten. Der Wert 1 bedeutet folglich, dass die Aussagen als erfunden eingestuft wurden,

¹³ Ein Kolmogorov-Smirnov-Test deutet darauf hin, dass die Variable „Glaubwürdigkeit“ in den beiden Experimentalgruppen normalverteilt ist (Vermeidung Urteil erfunden: $p = .17$; Vermeidung Urteil wahr: $p = .06$). Der Levene-Test auf Varianzhomogenität deutet darauf hin, dass die Varianzen der beiden Experimentalgruppen gleich sind, $F(1, 35) = 0.25$, $p = .88$.

Der Wert 2, dass die Aussagen als wahr eingestuft wurden. Insgesamt stuften 62.2% der Probanden die Aussagen als erfunden ein. Um zu testen, ob Probanden in der Bedingung „Vermeidung Urteil erfunden“ häufiger die Aussagen als wahr einstufen und Probanden in der Bedingung „Vermeidung Urteil wahr“ die Aussagen häufiger als erfunden einstufen, wurde ein Pearsons Chi-Quadrat Test durchgeführt. Alle erwarteten Werte waren größer als 5. Die Aussagen wurden von 83.3% ($N = 15$) der Probanden der Bedingung „Vermeidung Urteil wahr“ (im Vergleich 42.1% der Probanden der Bedingung „Vermeidung Urteil erfunden“) als erfunden eingestuft. Dagegen wurden die Aussagen von 57.9% ($N = 11$) der Probanden der Bedingung „Vermeidung Urteil erfunden“ (im Vergleich zu 16.7% der Probanden der Bedingung „Vermeidung Urteil wahr“) als wahr eingestuft. Die Ergebnisse deuten darauf hin, dass es einen signifikanten Zusammenhang zwischen der Urteilsvermeidung einerseits und dem Umstand, dass Probanden die Aussage als wahr oder erfunden einstufen, gibt, $\chi^2(1) = 6.68$, $p = .01$. Dies bedeutet, dass – basierend auf der Odds Ratio (relative Chance, Fahrmeier, Künstler, Pigeot, & Tutz, 2001) – die Chance, die Aussagen als wahr einzustufen, 6.88mal höher war, wenn Probanden der Bedingung „Vermeidung Urteil erfunden“ zugehörten (cf. Field, 2009, S. 699f.). Dieser Befund stimmt mit den Ergebnissen im Hinblick auf das metrische Glaubwürdigkeitsurteil überein.

3.3.3 *Argumente allgemein*

Um die Anzahl der Argumente, welche die Probanden insgesamt angaben, zu berechnen, wurden die Argumente (welche die Probanden durchnummeriert hatten) aufsummiert. Im Durchschnitt gaben Probanden 2.54

($SD = 1.46$) Argumente für ihr Urteil ab. Es gab keinen signifikanten Effekt der Urteilsvermeidung auf die Anzahl der Argumente, $F < 1$.¹⁴ Somit gaben die Probanden der Bedingung „Vermeidung Urteil erfunden“ genauso viele Argumente ($M = 2.53$, $SD = 1.12$) wie Probanden der Bedingung „Vermeidung Urteil wahr“ ($M = 2.56$, $SD = 1.79$) an.

3.3.4 Argumente für Urteil „wahr“ vs. „erfunden“

Um zu untersuchen, ob Probanden das Interview verzerrt, d.h. im Einklang mit ihrem Urteilsziel verarbeiteten, wurde die Anzahl der Argumente berechnet, die Probanden zur Begründung der Schuld bzw. Unschuld des Schülers angaben. Zwei voneinander unabhängige Rater kategorisierten die Argumente als „Argument wahr“ oder „Argument erfunden“. Wenn Probanden z.B. ihr Urteil rechtfertigten, indem sie argumentierten, dass „der Schüler eine Menge Details nennt, was sehr überzeugend ist“ (Argument 1), aber dass „er einige Informationen mehrmals wiederholt, was ihn unglaublich erscheinen lässt“ (Argument 2), so hätten die Probanden ein „Argument wahr“ (siehe Argument 1) und ein „Argument erfunden“ (siehe Argument 2) abgegeben. Sowohl für „Argument wahr“ als auch für „Argument erfunden“ stimmten die beiden Rater zu 90.9% überein („Argument wahr“: Cohens $\kappa = .87$, $p < .001$; „Argument erfunden“: Cohens $\kappa = .88$, $p < .001$).

¹⁴ Ein Kolmogorov-Smirnov-Test deutet darauf hin, dass die Variable „Argumente allgemein“ in den beiden Experimentalgruppen normalverteilt ist (Vermeidung Urteil erfunden: $p = .07$; Vermeidung Urteil wahr erfunden: $p = .13$). Der Levene-Test auf Varianzhomogenität deutet darauf hin, dass die Varianzen der beiden Experimentalgruppen gleich sind, $F(1, 35) = 2.60$, $p = .12$.

Eine ANOVA¹⁵ mit Messwiederholung mit „Art des Arguments“ (wahr vs. erfunden) als Innersubjektfaktor und „Urteilsvermeidung“ als Zwischensubjektfaktor zeigte einen signifikanten Haupteffekt der „Art des Arguments“ hin, $F(1, 35) = 4.18, p = .048, \eta^2 = .107$. Somit wurden mehr „Argumente erfunden“ ($M = 1.95, SD = 1.76$) als „Argumente wahr“ ($M = 1.08, SD = 1.36$) abgegeben. Erwartungsgemäß wurde dieser Haupteffekt durch eine signifikante Interaktion zwischen „Art des Arguments“ und „Urteilsvermeidung“ qualifiziert, $F(1, 35) = 4.69, p = .037, \eta^2 = .118$ (siehe Abbildung 2).

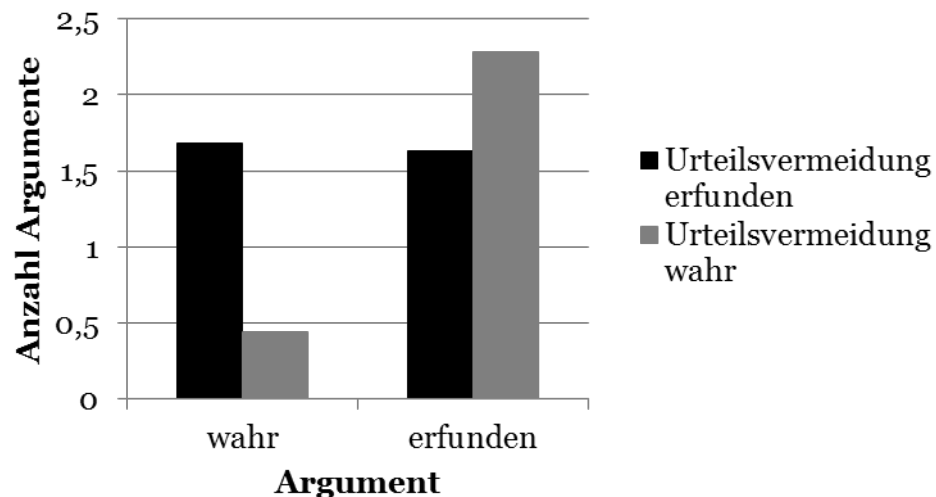


Abbildung 2. Anzahl angegebener Argumente in Abhängigkeit von der Valenz des Argumentes (wahr vs. erfunden) und der Urteilsvermeidung (wahr vs. erfunden) in Pilotstudie 2.

¹⁵ Ein Kolmogorov-Smirnov-Test deutet darauf hin, dass die Variablen „Argument wahr“ und „Argument erfunden“ in den beiden Experimentalgruppen nicht normalverteilt sind (Argument wahr: $ps < .005$; Argument erfunden: $ps < .04$). Der Levene-Test auf Varianzhomogenität deutet darauf hin, dass die Varianzen der beiden Experimentalgruppen für „Argument wahr“ nicht gleich sind, $F(1, 35) = 6.66, p = .01$, für „Argument erfunden“ jedoch gleich sind, $F(1, 35) = 0.96, p = .33$.

Dieser Befund deutet darauf hin, dass Probanden mehr „Argumente wahr“ in der Bedingung „Vermeidung Urteil erfunden“ ($M = 1.68$, $SD = 1.42$) als in der Bedingung „Vermeidung Urteil wahr“ ($M = 0.44$, $SD = 0.98$) angaben und dass Probanden mehr „Argumente erfunden“ in der Bedingung „Vermeidung Urteil wahr“ ($M = 2.28$, $SD = 2.08$) als in der Bedingung „Vermeidung Urteil erfunden“ ($M = 1.63$, $SD = 1.38$) angaben. Einfache Haupteffekt-Analysen für Messwiederholungen deuten darauf hin, dass die zwei Experimentalgruppen sich signifikant für „Argument wahr“ unterschieden, $F(1, 35) = 9.46$, $p = .004$, $\eta^2 = .21$, aber nicht für „Argument erfunden“, $p = .271$.

3.4 Diskussion

Die Ergebnisse zeigen, dass die Urteilszielmanipulation erfolgreich war. So tendierten die Probanden mit der Bedingung „Vermeidung Urteil erfunden“ dazu, die Aussagen des Schülers als wahr einzustufen, wohingegen die Probanden mit der Bedingung „Vermeidung Urteil wahr“ dazu tendierten, die Aussagen als erfunden einzustufen (Hypothese 1). Probanden scheinen das Ziel des geringsten Aufwandes verfolgt zu haben und wollten durch ihr Urteil vermeiden, den zweiseitigen Fallbericht zu schreiben. Aber haben Probanden lediglich ihr Bedürfnis nach minimalem Aufwand befriedigt oder haben sie die Aussagen des Schülers auch im Einklang mit ihrem Urteilsziel verarbeitet? Eine gängige Methode, um zu erfassen, wie Informationen verarbeitet wurden (z.B. Giner-Sorolila & Chaiken, 1997; siehe auch Kapitel 1.6.2), ist das Erfragen von Gedanken oder Argumenten für ein Urteil. In Pilotstudie 2 konnte gezeigt werden, dass Probanden Glaubwürdigkeitsurteile im Einklang mit ihrem Urteilsziel geben. Im Allgemeinen gaben Probanden in beiden

Experimentalbedingungen gleich viele Argumente für ihr Urteil an. Des Weiteren deuten die Ergebnisse darauf hin, dass – im Allgemeinen – mehr Argumente angegeben wurden, welche die Schuld des Schülers unterstützen, als Argumente, welche die Unschuld des Schülers unterstützen. Interessanterweise gaben Probanden in der Bedingung „Vermeidung Urteil erfunden“ mehr „Argumente wahr“ als in der Bedingung „Vermeidung Urteil wahr“, was darauf hin deutet, dass die Probanden das Interview unterschiedlich und im Einklang mit ihrem Urteilsziel verarbeiteten. Allerdings trifft dies nicht auf „Argumente erfunden“ zu; hier unterschied sich die Anzahl der genannten Argumente in den beiden Experimentalgruppen nicht. Somit konnte Hypothese 2 teilweise bestätigt werden.

4 Experiment 1: Einfluss von unbewussten Urteilszielen und Verarbeitungsmodus auf Glaubwürdigkeitsurteile

4.1 Fragestellung und Hypothese

Pilotstudie 2 demonstrierte, dass das Urteilsziel der Probanden nicht nur das Glaubwürdigkeitsurteil, sondern auch die Verarbeitung urteilsrelevanter Informationen beeinflusst. Das erste Experiment dieser Dissertation sollte darüber hinaus den Effekt von Verarbeitungsmodus auf den Prozess der Urteilsbildung untersuchen und überprüfen, ob *Motivation* die treibende Kraft hinter dem Effekt der Urteilszielmanipulation aus Pilotstudie 2 ist. Dabei wurde angenommen, dass urteilszielinkonsistente Informationen bedrohlich wirken und eher zu einer systematischen Verarbeitung von Informationen führen (vgl. Giner-Sorolila & Chaiken, 1997). Wie bereits in Pilotstudie 2 wurde auch bei Experiment 1 angenommen, dass die Probanden der Bedingung „Vermeidung Urteil erfunden“ Aussagen eher als wahr einstufen, wohingegen die Probanden der Bedingung „Vermeidung Urteil wahr“ Aussagen eher als erfunden beurteilen sollten (Hypothese 1). Des Weiteren wurde angenommen, dass Probanden urteilsrelevante Informationen im Sinne des Urteilsziels verarbeiten, d.h. die Probanden der Bedingung „Vermeidung Urteil erfunden“ mehr Argumente angeben sollten, welche die Unschuld der Zielperson unterstützen, und die Probanden der Bedingung „Vermeidung Urteil wahr“ mehr Argumente liefern sollten, welche die Schuld der Zielperson stützen (Hypothese 2). Darüber hinaus bestand die Annahme, dass Probanden, die urteilszielinkonsistente Hinweise erhalten, im Allgemeinen mehr Argumente nennen als Probanden, denen

urteilsziel-konsistente Hinweisreize präsentiert werden (Hypothese 3; Giner-Sorolila & Chaiken, 1997).

4.2 Methode

4.2.1 Versuchsaufbau

Die Datenerhebung erfolgte computerbasiert (Unipark, www.unipark.info). Zunächst erhielten die teilnehmenden Lehramtsstudierenden allgemeine Informationen über die Studie (z.B. über Dauer und Aufwandsentschädigung) und wurden gebeten, Fragen hinsichtlich ihres Alters, Geschlechts, Studiengangs und Semesters zu beantworten. Anschließend wurde das gleiche Szenario wie in Pilotstudie 2 vorgelegt. Danach folgte wie in Pilotstudie 2 die Manipulation des Urteilsziels (Vermeidung Urteil erfunden vs. Vermeidung Urteil wahr). Allerdings wurden diesmal die Instruktionen zusätzlich durch eine Abbildung verdeutlicht (siehe Anhang I.c). Nachfolgend wurde der entweder urteilszielkonsistente oder urteilszielinkonsistente Hinweisreiz präsentiert. Dieser wurde in Anlehnung an die Studie von Giner-Sorolila und Chaiken (1997) entwickelt und informierte die Probanden, dass eine kurze Voruntersuchung für die Studie, an der sie gerade teilnahmen, bereits einige Zeit zuvor an der selben Universität durchgeführt worden war und dass 68% der Lehramtsstudierenden, die teilgenommen hatten, die Aussagen des Schülers als wahr (Hinweisreiz „wahr“) bzw. erfunden (Hinweisreiz „erfunden“) eingestuft haben. Wichtig war dabei, dass die Probanden von Experiment 1 darüber informiert wurden, dass bei der kurzen Voruntersuchung die Probanden keinen zweiseitigen Fallbericht hatten verfassen müssen. Durch diese Information sollte der Konformitätsdruck (vgl.

Asch, 1956, zu conformity) reduziert werden.¹⁶ Somit sollten Probanden der Bedingung „Vermeidung Urteil erfunden“ den „Hinweisreiz wahr“ als konsistent und den „Hinweisreiz erfunden“ als inkonsistent mit ihrem Urteilsziel wahrnehmen. Probanden der Bedingung „Vermeidung Urteil wahr“ sollten dagegen den „Hinweisreiz erfunden“ als konsistent und den „Hinweisreiz wahr“ als inkonsistent ansehen. Ein Vortest mit Lehramtsstudierenden ($N = 20$) zeigte keinen signifikanten Haupteffekt des Hinweisreizes auf das Glaubwürdigkeitsurteil, $p = .29$, und auf die Anzahl an Argumenten, $p = .25$.¹⁷ Somit unterschieden sich die Probanden, die entweder den „Hinweisreiz wahr“ oder „Hinweisreiz erfunden“ erhielten und bei denen kein Urteilsziel experimentell variiert wurde, nicht signifikant hinsichtlich ihres

¹⁶ Konformität kann definiert werden als die Tendenz, die eigene Wahrnehmung, Meinung oder das Verhalten so zu verändern, dass es konsistent mit der Gruppennorm ist (eigene Übersetzung; Brehm, Kassir, & Fein, 2002, S. 222). „Gruppe“ wäre in diesem Fall die Gruppe der Lehramtsstudierenden, „Norm“ wäre die Tendenz, die Aussagen als wahr bzw. erfunden einzustufen. Um von der Gruppennorm abweichen zu können, brauchten die Probanden eine überzeugende Rechtfertigung. Die „Tatsache“, dass die Probanden im kurzen Vortest nicht die Konsequenzen für ihr Urteil zu tragen hatten, d.h. keinen zweiseitigen Fallbericht schreiben mussten, sollte als Rechtfertigung für ein abweichendes Urteil angesehen werden.

¹⁷ Ein Kolmogorov-Smirnov-Test deutet darauf hin, dass die Variablen „Glaubwürdigkeit“ (Hinweisreiz wahr: $p = .20$; Hinweisreiz erfunden: $p = .20$) und „Argumente allgemein“ (Hinweisreiz wahr: $p = .20$; Hinweisreiz erfunden: $p = .04$) in mindestens einer der Experimentalgruppen nicht normalverteilt ist. Der Levene-Test auf Varianzhomogenität deutet darauf hin, dass die Varianzen der beiden Experimentalgruppen jeweils gleich sind (Glaubwürdigkeit: $F(1, 18) = 0.04$, $p = .95$; Argumente allgemein: $F(1, 18) = 1.85$, $p = .19$).

Glaubwürdigkeitsurteils und der Anzahl an Argumenten. Dies deutet darauf hin, dass die Probanden nicht per se durch den Inhalt des Hinweisreizes im Hinblick auf ihr Glaubwürdigkeitsurteil beeinflusst wurden und dass sie sich nicht bedroht fühlten oder motiviert waren, ihren Verarbeitungsaufwand zu erhöhen, nachdem sie darüber informiert wurden, wie andere Personen derselben Gruppe (selbe Universität und Lehramtsstudierende) die Aussagen des Schülers wahrgenommen haben. Somit kann davon ausgegangen werden, dass durch den Hinweisreiz kein Konformitätsdruck ausgelöst wurde. Nach der Präsentation des Hinweisreizes folgte das Interview aus Pilotstudie 1 und 2, in dem der Kollege den verdächtigten Schüler befragt. Anschließend wurden die Probanden gebeten, ihr Glaubwürdigkeitsurteil abzugeben und ihr Urteil zu begründen. Wie zuvor wurden auch hier die Probanden gebeten, ihre Argumente durchzunummerieren; dieses Mal wurden sie jedoch auch gebeten, ihre Argumente hinsichtlich ihrer Valenz zu kategorisieren (Argument wahr vs. Argument erfunden). Zum Schluss wurde den Probanden für ihre Teilnahme gedankt und sie wurden gebeten – falls erforderlich – den zweiseitigen Fallbericht zu schreiben. Sofern die Probanden den Fallbericht verfassen sollten, wurden sie gebeten, sich beim Versuchsleiter zu melden, um weitere Informationen über den Fallbericht zu erhalten. Dieses Vorgehen wurde eingeführt, um sicherzustellen, dass alle Probanden, die den Fallbericht verfassen sollten, diesen auch tatsächlich verfassten. Zum Schluss wurden die Probanden über das tatsächliche Ziel der Studie aufgeklärt.

4.2.2 *Versuchsmaterial*

Das Versuchsmaterial war dasselbe wie in Pilotstudien 1 und 2.

4.2.3 *Versuchsdesign*

Experiment 1 folgte einem zweifaktoriellen Design mit dem Faktor „Urteilsvermeidung“ mit den Faktorstufen „Vermeidung Urteil erfunden“ und „Vermeidung Urteil wahr“ und dem Faktor „Hinweisreiz“ mit den Faktorstufen „Hinweisreiz wahr“ und „Hinweisreiz erfunden“.

4.2.4 *Stichprobe*

Insgesamt nahmen 46 Lehramtsstudierende an Experiment 1 teil. Alle Probanden verfassten den Fallbericht, sofern sie dies sollten¹⁸; somit mussten keine Probanden von den Analysen ausgeschlossen werden. Mehr als die Hälfte der Probanden war weiblich (73.9%), was für Lehramtsstudierende als gewöhnliche Verteilung angesehen werden kann. Im Durchschnitt waren die Probanden 22.89 ($SD = 2.33$) Jahre alt und studierten im 4.33 ($SD = 2.34$) Semester.

4.3 Ergebnisse

4.3.1 *Glaubwürdigkeitsurteil*

Im Durchschnitt wurden die Aussagen des Schülers als $M = 5.74$ ($SD = 1.90$) glaubwürdig eingestuft. Dieser Wert unterschied sich nicht signifikant

¹⁸ Dass alle Probanden den Instruktionen folgten und den Fallbericht verfassten, kann darauf zurückgeführt werden, dass der Versuchsleiter die Probanden zu Beginn des Experiments darauf hinwies, dass unter bestimmten Bedingungen ein zweiseitiger Fallbericht verfasst werden sollte.

vom Skalenmittelwert 5.5, $p = .40$. Eine ANOVA¹⁹ mit dem Glaubwürdigkeitsurteil als abhängige Variable und „Urteilsvermeidung“ und „Hinweisreiz“ als unabhängige Variablen zeigte keinen Effekt, alle $ps > .20$.

4.3.2 *Binäres Glaubwürdigkeitsurteil*

Um zu überprüfen, ob Probanden in Abhängigkeit ihres Urteilsziels dazu tendieren, ein bestimmtes Urteil abzugeben, wurde das metrische Glaubwürdigkeitsurteil wie in Pilotstudie 2 in ein binäres Glaubwürdigkeitsurteil transformiert. Insgesamt stuften 54.3% der Probanden die Aussagen als erfunden ein. Um zu testen, ob Probanden in der Bedingung „Vermeidung Urteil erfunden“ häufiger die Aussagen als wahr einstufen und Probanden in der Bedingung „Vermeidung Urteil wahr“ häufiger die Aussagen als erfunden einstufen, wurde ein Pearsons Chi-Quadrat Test durchgeführt. Alle erwarteten Werte waren größer als 5. Die Aussagen wurden von 76.2% ($N = 16$) der Probanden der Bedingung „Vermeidung Urteil wahr“ (im Vergleich 23.8% der Probanden der Bedingung „Vermeidung Urteil erfunden“) als erfunden eingestuft. Dagegen wurden die Aussagen von 56.0% ($N = 14$) der Probanden der Bedingung „Vermeidung Urteil erfunden“ (im Vergleich zu 44.0% der Probanden der Bedingung „Vermeidung Urteil wahr“) als wahr eingestuft. Die Ergebnisse deuten darauf hin, dass es einen signifikanten Zusammenhang zwischen den Faktoren „Urteilsvermeidung“ und dem Fakt,

¹⁹ Ein Kolmogorov-Smirnov-Test deutet darauf hin, dass die Variablen „Glaubwürdigkeit“ in mindestens einer der Experimentalgruppen nicht normalverteilt ist, $ps < .14$. Der Levene-Test auf Varianzhomogenität deutet darauf hin, dass die Varianzen der Experimentalgruppen nicht gleich sind ($ps < .03$).

dass Probanden die Aussagen als wahr oder erfunden eingestuft haben, gibt, $\chi^2(1) = 4.88, p = .027$. Dies bedeutet, dass – basierend auf der Odds Ratio (relative Chance, Fahrmeir, Pigeot, Künstler, & Tutz, 2009) – die Chance, die Aussagen als wahr einzustufen, für Probanden der Bedingung „Vermeidung Urteil erfunden“ im Vergleich zur Bedingung „Vermeidung Urteil wahr“ 4.07mal höher war (cf. Field, 2009, S. 699 f.). Es gab keinen signifikanten Zusammenhang zwischen „Hinweisreiz“ und dem Urteil der Probanden, $\chi^2(1) = 0.32, p = .57$.

4.3.3 Argumente allgemein

Um die Anzahl der Argumente, welche die Probanden insgesamt angaben, zu berechnen, wurden die Argumente (welche die Probanden durchnummeriert und kategorisiert hatten) aufsummiert. Im Durchschnitt gaben Probanden 3.54 ($SD = 1.97$) Argumente für ihr Urteil an. Eine ANOVA²⁰ mit der Anzahl der Argumente als abhängige Variable und „Urteilsvermeidung“ und „Hinweisreiz“ als unabhängige Variablen ergab einen marginal signifikanten Haupteffekt des Hinweisreizes, $F(1, 42) = 3.87, p = .056, \eta^2 = 0.84$. Dieser Effekt wurde durch eine signifikante Interaktion zwischen Urteilsvermeidung und Hinweisreiz qualifiziert (siehe Abbildung 3), $F(1, 42) = 9.93, p = .003, \eta^2 = 0.19$. Kontrastanalysen (berechnet nach der Formel von

²⁰ Ein Kolmogorov-Smirnov-Test deutet darauf hin, dass die Variable „Argumente allgemein“ in mindestens einer der Experimentalgruppen nicht normalverteilt ist, $ps > .001$. Der Levene-Test auf Varianzhomogenität deutet darauf hin, dass die Varianzen der Experimentalgruppen jeweils gleich sind (Urteilsvermeidung: $F(1, 44) = 0.28, p = .60$; Hinweisreiz: $F(1, 44) = 2.32, p = .63$).

Winer, 1971) deuten darauf hin, dass in der Bedingung „Vermeidung Urteil wahr“ Probanden signifikant mehr Argumente angaben, nachdem sie den urteilszielinkonsistenten „Hinweisreiz wahr“ ($M = 4.64$, $SD = 2.13$) im Vergleich zum urteilszielkonsistenten „Hinweisreiz erfunden“ ($M = 2.00$, $SD = 1.15$) erhielten, $p = .003$. In der Bedingung „Vermeidung Urteil erfunden“ gaben die Probanden ebenfalls signifikant mehr Argumente an, nachdem sie den urteilszielinkonsistenten „Hinweisreiz erfunden“ ($M = 4.11$, $SD = 1.36$) im Vergleich zum urteilszielkonsistenten „Hinweisreiz wahr“ ($M = 3.50$, $SD = 1.96$) erhielten, $p = .003$.

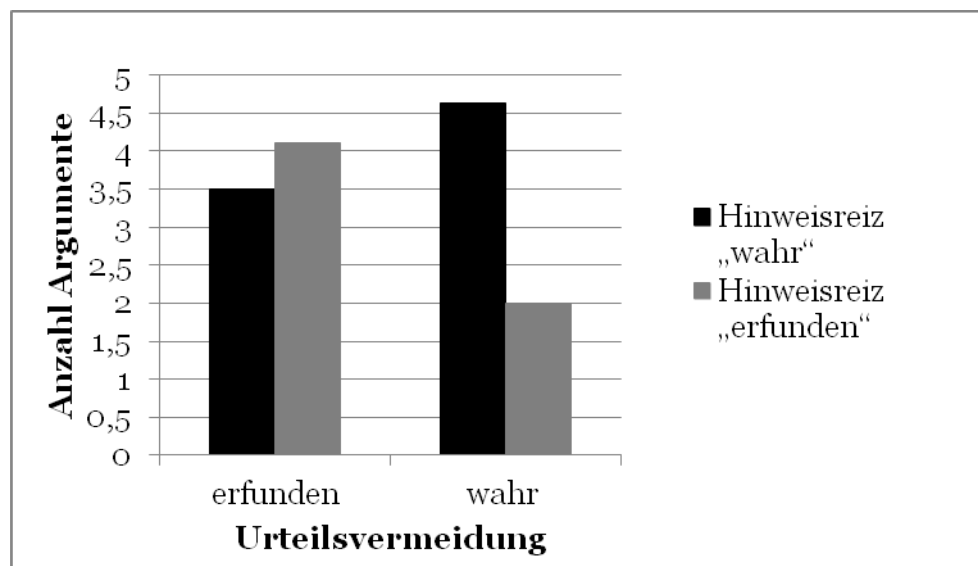


Abbildung 3. Anzahl angegebener Argumente in Abhängigkeit von der Urteilsvermeidung (wahr vs. erfunden) und des Hinweisreizes (wahr vs. erfunden) in Experiment 1.

4.3.4 Argumente für Urteil „wahr“ vs. „erfunden“

Um zu untersuchen, ob Probanden das Interview unterschiedlich und im Einklang mit ihrem Urteilsziel verarbeiteten, wurde die Anzahl der Argumente aufsummiert, welche die Probanden zur Begründung der Schuld bzw. Unschuld

des Schülers angaben. Eine ANOVA mit Messwiederholung²¹ mit „Art des Arguments“ (wahr vs. erfunden) als Innersubjektfaktor und „Urteilsvermeidung“ als Zwischensubjektfaktor deutete auf einen signifikanten Haupteffekt des Faktors „Art des Arguments“ hin, $F(1, 42) = 10.68$, $p = .002$, $\eta^2 = 0.20$. Erwartungsgemäß wurde dieser Haupteffekt durch eine signifikante Interaktion zwischen den Faktoren „Art des Arguments“ und „Urteilsvermeidung“ qualifiziert (siehe Abbildung 4), $F(1, 42) = 6.67$, $p = .013$, $\eta^2 = 0.14$. Dieser Befund deutet darauf hin, dass Probanden mehr „Argumente wahr“ in der Bedingung „Vermeidung Urteil erfunden“ ($M = 2.58$, $SD = 1.43$) als in der Bedingung „Vermeidung Urteil wahr“ ($M = 1.78$, $SD = 1.37$) angaben und dass Probanden mehr „Argumente erfunden“ in der Bedingung „Vermeidung Urteil wahr“ ($M = 1.59$, $SD = 1.39$) als in der Bedingung „Vermeidung Urteil erfunden“ ($M = 1.21$, $SD = 0.71$) angaben. Einfache Haupteffekt-Analysen für Messwiederholungen deuten darauf hin, dass die zwei Experimentalgruppen sich signifikant für „Argument wahr“, $p = .003$, und für „Argument erfunden“, $p = .003$, unterscheiden.

²¹ Ein Kolmogorov-Smirnov-Test deutet darauf hin, dass die Variablen „Argumente wahr“ und „Argumente erfunden“ in mindestens einer der Experimentalgruppen nicht normalverteilt ist, $ps > .001$. Der Levene-Test auf Varianzhomogenität deutet darauf hin, dass die Varianzen der Experimentalgruppen jeweils gleich sind (Urteilsvermeidung, Argument wahr: $F(1, 44) = 0.02$, $p = .90$; Urteilsvermeidung, Argument erfunden: $F(1, 44) = 2.38$, $p = .13$; Hinweisreiz, Argument wahr: $F(1, 44) = 0.41$, $p = .53$; Hinweisreiz, Argument erfunden: $F(1, 44) = 2.20$, $p = .15$).

Glaubwürdigkeitsurteile

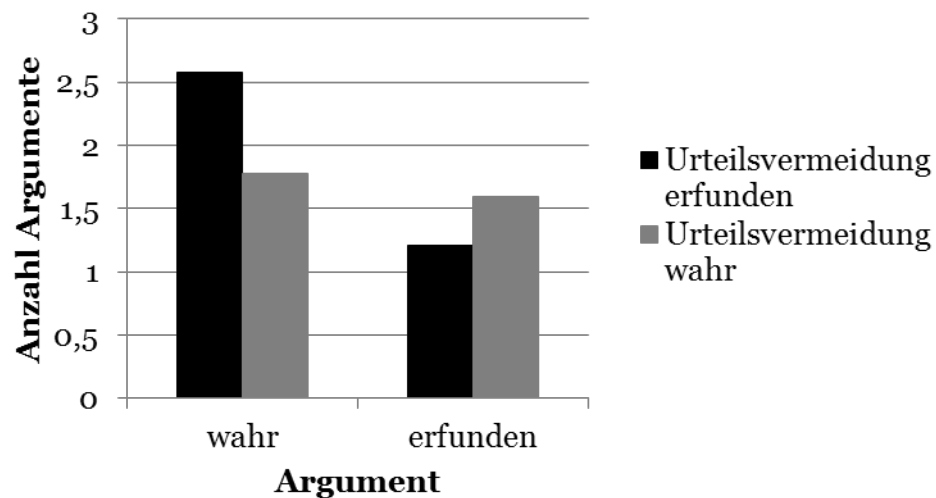


Abbildung 4. Anzahl angegebener Argumente in Abhängigkeit von der Valenz des Argumentes (wahr vs. erfunden) und der Urteilsvermeidung (wahr vs. erfunden) in Experiment 1.

4.4 Diskussion

Im Durchschnitt beurteilten die Probanden die Aussagen des Schülers als durchschnittlich glaubwürdig. Die Urteilszielmanipulation kann als erfolgreich angesehen werden: Probanden in der Bedingung „Vermeidung Urteil erfunden“ beurteilten die Aussagen signifikant häufiger als wahr und Probanden in der Bedingung „Vermeidung Urteil wahr“ beurteilten sie signifikant häufiger als erfunden (Hypothese 1). Des Weiteren gaben Probanden in der Bedingung „Vermeidung Urteil erfunden“ mehr Argumente an, welche die Begründung der Unschuld unterstützten, als in der Bedingung „Vermeidung Urteil wahr“ und mehr Argumente für die Schuld des Schülers in der Bedingung „Vermeidung Urteil wahr“ im Vergleich zur Bedingung „Vermeidung Urteil erfunden“ (Hypothese 2). Dies deutet darauf hin, dass die Probanden insofern durch das Urteilsziel beeinflusst wurden, als dass sie nicht nur ein urteilszielentsprechendes Urteil abgaben, sondern auch die Aussagen auf eine in Richtung des Urteilsziels verzerrte Art und Weise verarbeiteten. Der

Hinweisreiz schien das binäre Glaubwürdigkeitsurteil nicht zu beeinflussen. Allerdings schien der Hinweisreiz die Art der Informationsverarbeitung beeinflusst zu haben. In der Bedingung „Vermeidung Urteil wahr“ gaben die Probanden signifikant mehr Argumente an, nachdem sie den urteilszielinkonsistenten „Hinweisreiz wahr“ im Vergleich zum urteilszielkonsistenten „Hinweisreiz erfunden“ erhalten hatten. In der Bedingung „Vermeidung Urteil erfunden“ gaben Probanden ebenfalls mehr Argumente nach Erhalt des inkonsistenten Hinweisreizes an (Hypothese 3). Dies deutet darauf hin, dass der Hinweisreiz als bedrohlich wahrgenommen wurde und die Probanden motivierte, mehr Argumente für ihr urteilszielentsprechendes Urteil zu generieren.

Zusammengefasst kann man sagen, dass Pilotstudie 2 und Experiment 1 demonstrierten, dass das Urteilsziel das Glaubwürdigkeitsurteil und die Verarbeitung urteilsrelevanter Informationen beeinflusst. Experiment 1 konnte darüber hinaus Hinweise liefern, dass der zugrunde liegende Mechanismus motivationaler Natur ist. Daher ist es auch angemessen davon auszugehen, dass Lehramtsstudierende einem Urteilsziel folgen (vgl. Definition von Ziel und Motivation oben), nämlich dem Ziel, ihren Aufwand zu minimieren. Des Weiteren konnte gezeigt werden, wie der Verarbeitungsmodus (variiert über die Hinweisreize) die Verarbeitung von urteilsrelevanten Informationen beeinflusst.

Das Stimulusmaterial, das in Pilotstudie 1 und 2 und Experiment 1 verwendet wurde, kann als künstliches Material angesehen werden, das üblicherweise in Studien verwendet wird, bei denen Urteilen unterliegende

Prozesse untersucht werden (vgl. Reinhard, 2010; Reinhard & Sporer, 2008, 2010). Des Weiteren sind Lehramtsstudierende, die sowohl in Pilotstudie 1 und 2 als auch Experiment 1 Probanden waren, noch nicht im Berufsleben integriert und haben daher nur wenige oder gar keine Erfahrungen im Umgang mit Schülern. Laut den Ergebnissen von Reinhard, Dickhäuser, et al. (2011), haben jedoch Lehrkräfte, die bereits im Berufsleben stehen, eine stärkere Tendenz, Aussagen von Schülern als wahr einzustufen. Daher ist es interessant zu untersuchen, ob die Motivation, Aufwand zu minimieren, das Glaubwürdigkeitsurteil auch in einer Situation beeinflusst, in der den Probanden Videomaterial mit Schülern vorgelegt wird, die tatsächlich die Wahrheit sagen oder lügen. Interessant ist dabei auch, dass Videomaterial neben verbalen Informationen – die auch in Transkripten enthalten sind – auch nonverbale Informationen enthält. Somit stellt sich die Frage, ob ein Hinweisreiz, wie er in Experiment 1 verwendet wurde, auch den Verarbeitungsmodus und somit die Informationsverarbeitung in realen Lügendetektionssituationen von im Berufsleben stehenden Lehrkräften beeinflusst. Da erfahrene Lehrkräfte eine ausgeprägte Wahrheitsverzerrung haben (vgl. Reinhard, Dickhäuser, et al., 2011), stellt sich weiterhin die Frage, ob diese natürliche Wahrheitsverzerrung den Einfluss von Urteilsziel und Verarbeitungsmodus beeinträchtigt. Weiterhin wurde in Experiment 1 (und Pilotstudie 1 und 2) ein Mobbing-Vorfall beurteilt. Daher ist es darüber hinaus interessant zu untersuchen, ob der Effekt von Urteilsziel und Verarbeitungsmodus über verschiedene Schulszenarien hinweg robust auftritt.

5 Experiment 2: Einfluss von unbewussten Urteilszielen und Verarbeitungsmodus auf die Akkuratheit von Glaubwürdigkeitsurteilen

5.1 Fragestellung und Hypothesen

Wie bereits oben erläutert wurde, ist bei der Erforschung von Glaubwürdigkeitsurteilen immer auch die Akkuratheit dieser Urteile von Interesse. Um die Akkuratheit von Glaubwürdigkeitsurteilen untersuchen zu können, ist es notwendig, die „Wahrheit“ über das zu beurteilende oder das zu untersuchende Objekt zu kennen (West & Kenny, 2011), d.h. es ist erforderlich zu wissen, ob die zu beurteilende Person die Wahrheit sagt oder lügt. Nur dann ist es möglich zu bestimmen, inwieweit sich das Urteil über die Glaubwürdigkeit einer Person vom tatsächlichen Wahrheitsstatus unterscheidet. Somit kann die Akkuratheit eines Glaubwürdigkeitsurteils nur bestimmt werden, wenn bekannt ist, ob die Aussagen der zu beurteilenden Person wahr oder erfunden sind. In Pilotstudie 1 und 2 und Experiment 1 wurden die Probanden gebeten, die Glaubwürdigkeit der Aussagen eines Schülers während eines Interviews zu beurteilen. Dieses Interview war – obwohl es auf einer tatsächlichen Aussage basierte – konstruiert; somit existierte insoweit keine „Wahrheit“. Da der Fokus in Experiment 2²² auf der Akkuratheit von Glaubwürdigkeitsurteilen liegt, wurden in Experiment 2 tatsächliche Aussagen von Schülerinnen und Schülern

²² Die Daten zu Experiment 2 sowie Experiment 3 wurden für ein Manuskript verwendet, welches zur Publikation in einer Fachzeitschrift vorbereitet wird (Marksteiner, Plieninger, Reinhard, & Dickhäuser, 2013).

zur Beurteilung eingesetzt. Im Gegensatz zu den vorangegangenen Experimenten lag hier das Interesse nicht auf der Erforschung des Prozesses, der verzerrten Glaubwürdigkeitsurteilen unterliegt, sondern auf dem Effekt, den Urteilsziel und Verarbeitungsmodus auf die Fähigkeit haben, akkurat zwischen wahren und erfundenen Aussagen differenzieren zu können. Zu diesem Zweck wurde die Manipulation des Urteilsziels, wie sie in Pilotstudie 2 und Experiment 1 verwendet wurde, an ein häufig verwendetes experimentelles Design zur Untersuchung des Erkennens von Lüge und Wahrheit angepasst, bei dem keine schriftlichen Aussagen, sondern mehrere, auf Video aufgenommene Interviews präsentiert werden (vgl. C. F. Bond & DePaulo, 2006). Des Weiteren nahmen an Experiment 2 Lehrkräfte teil, die sich zu diesem Zeitpunkt bereits im Schuldienst befanden. Im Gegensatz zu den vorangegangenen Studien wurde ein Szenario gewählt, bei dem die Stimuluspersonen befragt wurden, ob sie bei einem Test unerlaubte Hilfsmittel verwendet hätten. Des Weiteren handelte es sich bei dem Stimulusmaterial um Material, anhand dessen überprüft werden konnte, wie akkurat ein Proband die Aussagen der Stimulusperson als wahr oder erfunden einstufen kann.

Wie auch in den vorangegangenen Experimenten, wurde in Experiment 2 angenommen, dass das Urteilsziel das Glaubwürdigkeitsurteil beeinflusst. In Experiment 2 wurde das Glaubwürdigkeitsurteil wie üblich (vgl. C. F. Bond & DePaulo, 2006) anhand eines binären Items (1 = „Die Aussagen sind wahr.“, 2 = „Die Aussagen sind erfunden.“) erfasst. Es wurde davon ausgegangen, dass die Probanden in der Bedingung „Vermeidung Urteil erfunden“ häufiger die Aussagen eines Schülers als wahr einstufen, d.h. eine Wahrheitsverzerrung

zeigen (Hypothese 1a). Dagegen sollten Probanden der Bedingung „Vermeidung Urteil wahr“ häufiger das Urteil „Die Aussagen sind erfunden.“ wählen, d.h. eine Lügenverzerrung aufweisen (Hypothese 1b). Da bisherige Studien zeigen konnten, dass weder Laien (C. F. Bond & DePaulo, 2006) noch Lehrkräfte (Reinhard, Dickhäuser, et al., 2011) besser im Erkennen von wahren und erfunden Aussagen sind als das Zufallsniveau, wurde weiterhin angenommen, dass die Trefferquote der Lehrkräfte ebenfalls nicht über dem Zufallsniveau liegt (Hypothese 2). Da angenommen wird, dass Probanden der Bedingung „Vermeidung Urteil erfunden“ häufiger das Urteil „Die Aussagen sind wahr.“ abgeben, und da bisherige Befunde zeigen, dass Personen aufgrund ihrer Tendenz, Aussagen als wahr einzustufen, wahre Aussagen besser einstufen (Wahrheitseffekt, Levine et al., 1999), wurde davon ausgegangen, dass Probanden der Bedingung „Vermeidung Urteil erfunden“ eine höhere Trefferquote beim Erkennen von wahren im Vergleich zu erfundenen Aussagen aufweisen (Hypothese 3a). Erklärt werden kann dies durch die höhere Wahrscheinlichkeit, wahre Aussagen als solche zu erkennen, wenn mehr Urteile „wahr“ abgegeben werden. Daher wurde auch angenommen, dass Probanden der Bedingung „Vermeidung Urteil wahr“ erfundene Aussagen besser als solche identifizieren als wahre Aussagen als wahr (Hypothese 3b). Laut den Befunden von Masip et al. (2009) und Reinhard, Sporer, et al. (2011) konnte zudem davon ausgegangen werden, dass eine systematische (vs. heuristische) Verarbeitung zu einer höheren Akkuratheit führt. Dies wird dadurch erklärt, dass eine systematische Verarbeitung zur Nutzung verbaler Informationen führt (Reinhard & Sporer, 2008, 2010) und mehr verbale Hinweisreize als valide

Lügenindikatoren gelten (DePaulo et al., 2003). Somit wurde angenommen, dass eine systematische Verarbeitung – unabhängig vom Urteilsziel – zu einer höheren Akkuratheit führt (Hypothese 4a und 4b für die Bedingungen „Vermeidung Urteil erfunden“ und „Vermeidung Urteil wahr“). Dies kann dadurch begründet werden, dass eine verzerrte systematische Verarbeitung zwar eine selektive Verarbeitung von verbalen Informationen nach sich zieht (Chen & Chaiken, 1999), diese Informationen aber dennoch zum größten Teil valide sind (vgl. C. F. Bond & DePaulo, 2006).

5.2 Methode

5.2.1 Versuchsablauf

Die Erhebung der Daten erfolgte computerbasiert (Unipark, www.unipark.info). Den Probanden wurde per E-Mail ein Link zugeschickt. Zunächst erhielten die Probanden allgemeine Informationen über die Dauer der Studie, das Ziel der Untersuchung und die Aufwandsentschädigung. Außerdem wurde ihnen Anonymität und die ausschließliche Verwendung der Daten für wissenschaftliche Zwecke zugesichert. Anschließend schauten die Probanden eine kurze Videosequenz, in der ein Mitarbeiter des Lehrstuhls zu sehen war, der die Probanden bat, die Audioeinstellungen an ihrem Computer zu überprüfen.²³ Auf der folgenden Seite wurden die Probanden gebeten, Angaben bezüglich ihres Alters, Geschlechts, Berufs und ihrer Berufserfahrung (in Jahren und Monaten) zu machen. Danach wurde ihnen ein Szenario präsentiert,

²³ An dieser Stelle möchte ich Martin Scharmach für die Erlaubnis zur Benutzung der Videosequenz danken.

welches sich von den Szenarien der vorherigen Experimente unterscheidet. Sie wurden gebeten, sich vorzustellen, sie seien Lehrer einer Klasse, in der Schüler gemeinsam in Zweier-Gruppen einen Test bearbeitet hatten. Am Ende des Tests erhielt jede Zweier-Gruppe eine gemeinsame Gruppennote; es wurden keine Einzelnoten vergeben. Nach der Benotung aller Gruppen fiel auf, dass einige Schüler-Gruppen viel besser als erwartet abgeschnitten hatten. Der Lehrer nimmt nun an, dass diese Gruppen unerlaubte Hilfsmittel bei der Bearbeitung des Tests verwendet haben könnten und stellt den besagten Schülern in Einzelinterviews Fragen bezüglich der Lösung der Aufgaben und ob sie unerlaubte Hilfsmittel verwendet hätten. Nach der Präsentation des Szenarios erfolgte die Manipulation des Urteilsziels (Vermeidung Urteil erfunden vs. Vermeidung Urteil wahr), welche ähnlich der Manipulation war, die in den vergangenen beiden Experimenten genutzt wurde. Alle Probanden wurden informiert, dass acht Interviews zwischen dem Lehrer und jeweils einem der besagten Schüler folgen würden und dass sie gebeten werden würden, die Glaubwürdigkeit der Aussagen des jeweiligen Schülers anhand eines binären Items (1 = „Die Aussagen sind wahr.“, 2 = „Die Aussagen sind erfunden.“) zu beurteilen. Des Weiteren wurden sie gebeten, ihr Urteil kurz mit Hilfe weniger Stichworte zu begründen. Probanden der Bedingung „Vermeidung Urteil erfunden“ wurden gebeten, nach Beurteilung aller acht Interviews eine detailliertere, zusätzliche Begründung für alle Schüler abzugeben, deren Aussagen sie als erfunden eingestuft hatten. In der Bedingung „Vermeidung Urteil wahr“ wurden die Probanden ebenfalls gebeten detailliertere Begründungen abzugeben, allerdings für alle Schüleraussagen, die sie als wahr

eingestuft hatten. Wie in Experiment 1 wurden die Instruktionen anhand einer Graphik dargestellt, um sie für die Probanden klarer zu machen. Danach folgte der Instruktionscheck für die Urteilszielmanipulation und die Probanden wurden gefragt, unter welchen Umständen sie eine detailliertere, zusätzliche Begründung abgeben müssten. Die Antwortoptionen waren „... wenn ich denke, die Aussagen sind wahr“ und „... wenn ich denke, die Aussagen sind erfunden“; zusätzlich gab es die Option „Ich weiß es nicht.“. Sofern eine falsche Antwort gegeben wurde (oder die Option „Ich weiß es nicht.“ gewählt wurde), wurden die Probanden gebeten, die Instruktionen nochmals sorgfältig zu lesen. Danach wurden ihnen die Fragen zur Überprüfung des Instruktionsverständnisses erneut gestellt. Schließlich wurde ein entweder urteilszielkonsistenter oder -inkonsistenter Hinweisreiz präsentiert. Der Hinweisreiz war dem in Experiment 1 ähnlich und informierte die Probanden, dass eine kurze Voruntersuchung bereits einige Zeit vor der Studie durchgeführt worden war, bei dem Lehrkräfte ebenfalls gebeten wurden, die Glaubwürdigkeit der acht folgenden Schüleraussagen einzustufen. Weiterhin wurden die Ergebnisse der kurzen Voruntersuchung dargestellt, welche die Probanden informierte, dass sechs der acht Schüleraussagen entweder als wahr (Hinweisreiz wahr) oder erfunden (Hinweisreiz erfunden) beurteilt wurden. Es wurde davon ausgegangen, dass Probanden der Bedingung „Vermeidung Urteil erfunden“ den „Hinweisreiz erfunden“ als urteilszielinkonsistent und den „Hinweisreiz wahr“ als urteilszielkonsistent wahrnehmen. Dagegen sollten Probanden der Bedingung „Vermeidung Urteil wahr“ den „Hinweisreiz erfunden“ als urteilszielkonsistent und den „Hinweisreiz wahr“ als urteilszielinkonsistent wahrnehmen. Der Effekt

des Hinweisreizes auf die Urteilsverzerrung wurde vorgetestet ($N = 33$ Lehrkräfte); es zeigte sich kein signifikanter Effekt, $p = .516$. Somit unterschieden sich Probanden nicht signifikant hinsichtlich ihrer Urteilsverzerrung, nachdem sie einen der beiden Hinweisreize erhalten hatten ($M_{\text{WahrHinweis}} = 5.20$, $SD_{\text{WahrHinweis}} = 1.37$; $M_{\text{LügenHinweis}} = 4.83$, $SD_{\text{LügenHinweis}} = 1.76$). Dies deutet darauf hin, dass das Glaubwürdigkeitsurteil der Probanden nicht durch den Inhalt des Hinweisreizes beeinflusst wird und somit kein Konformitätsdruck (vgl. conformity; Asch, 1956) vorhanden war. Anschließend folgten die acht Interviews, die Probanden beurteilten die Glaubwürdigkeit der Schüleraussagen und begründeten die Aussagen kurz (kurze Urteilsbegründung). Wie zuvor wurden die Probanden auch diesmal gebeten, ihre Argumente zu nummerieren. Nach Beurteilung aller acht Schülerinterviews wurden die Probanden gebeten die detailliertere Begründung (zusätzliche Begründung) abzugeben, sofern sie ein entsprechendes Urteil abgegeben hatten. Am Ende wurde den Probanden für ihre Teilnahme gedankt, ihnen wurde die Möglichkeit gegeben, an der Verlosung von Gutscheinen eines namhaften Internetshops teilzunehmen und sie wurden über das tatsächliche Untersuchungsziel aufgeklärt.

5.2.2 Versuchsmaterial

Insgesamt wurden den Probanden in Experiment 2 Videoaufnahmen von 24 Stimuluspersonen vorgelegt. Zwölf Stimuluspersonen waren männlich, zwölf waren weiblich. Alle Stimuluspersonen waren Schüler einer Schule aus der Region und 18 Jahre alt. Der Raum, in dem das Interview durchgeführt wurde, war mit einer Videokamera ausgestattet. Der Torso und die Beine der

Stimuluspersonen waren auf den Videoaufnahmen sichtbar. Die Tonqualität war bei allen Videoaufnahmen gleich gut. Alle Stimuluspersonen wurden zu einem Test befragt, den sie kurz zuvor mit einem Partner bearbeitet hatten. Die Zweier-Gruppen bestanden jeweils aus einem Schüler und einem Strohmännchen gleichen Geschlechts. Der Test beinhaltete verschiedene Aufgaben, die ursprünglich aus einem deutschen Intelligenztest stammen (Berliner Intelligenz Struktur Test; Jäger, 2006). Der zu bearbeitende Test enthielt fünf Arten von Aufgaben, die innerhalb von zehn Minuten gelöst werden mussten. Allen Stimuluspersonen wurde zu Anfang mitgeteilt, dass sie an einer Studie zum Thema „Lernen im Klassenzimmer“ teilnehmen würden. Sie wurden zufällig einer von zwei Bedingungen zugeordnet. Zu Beginn der Studie wurden alle Stimuluspersonen ihrem Lehrer und einem angeblich weiteren Schüler (des gleichen Geschlechts) vorgestellt, mit dem sie gemeinsam einen Test bearbeiten sollten. Tatsächlich handelte es sich bei dem weiteren Schüler um einen Verbündeten des Versuchsleiters. Die Stimulusperson (sowie der Verbündete) erhielten zu Beginn der Studie 10€ als Belohnung für die Teilnahme. Beide Schüler wurden darüber in Kenntnis gesetzt, dass sie beide jeweils weitere 20€ erhalten würden, wenn sie eine gute Leistung bei dem zu bearbeitenden Test zeigen würden. Nach dem der Lehrer die Art von Aufgaben, die im Test bearbeitet werden mussten, erklärte und anhand von Beispielen mit den Schülern geübt hatte, bearbeiteten die Schüler gemeinsam den Test. Circa 5 Minuten nach Beginn der Testzeit verließ der Lehrer aufgrund eines angeblichen Telefonanrufs den Raum. In der Lügen-Bedingung überredete der Verbündete die Stimulusperson, durch Nutzen eines Taschenrechners und/oder

der Lösungen, die auf dem Schreibtisch der Lehrkraft lagen, zu betrügen. Stimuluspersonen der Wahrheits-Bedingung wurden hingegen nicht überredet, unerlaubte Hilfsmittel zu verwenden. Nach Bearbeitung des Tests wurden die beiden Schüler jeweils in einen zweiten Raum zum Interview über die Bearbeitung des Tests gebeten. Die Interviews wurden auf Video aufgezeichnet. Der Interviewer informierte zunächst beide Schüler, dass ihre Leistung bei dem Test sehr gut war und dass sie die weiteren 20€ als Belohnung erhalten würden. Die Schüler wurden sodann gebeten, ihre Strategien bei der Bearbeitung des Tests zu erläutern, und eine mögliche Erklärung für ihre sehr gute Leistung zu geben. Zum Schluss fragte der Interviewer, ob sie unerlaubte Hilfsmittel bei der Bearbeitung verwendet hatten. Diese Vorgehensweise wurde bereits von Exline, Thibaut, Hickey und Gumpert (1970) gewählt und unterscheidet sich von den meisten Vorgehensweisen zur Produktion von Stimulusmaterial in Studien zur Erforschung von Lügnerkennen (vgl. Vrij, 2008, S. 52). Anstatt die Stimuluspersonen mit einem gewissen Geldbetrag zu motivieren, den Interviewer von ihren Aussagen zu überzeugen, konnten sie eigenständig entscheiden, ob sie betrügen möchten und ob sie anschließend während des Interviews die Wahrheit sagen möchten oder lügen möchten. Somit lügen die Stimuluspersonen während des Interviews aus eigener Motivation heraus. Tatsächlich logen die Stimuluspersonen der Lügen-Bedingung während des Interviews. Ihre Motivation war neben dem möglichen Verlust der in Aussicht stehenden 20€ Belohnung wahrscheinlich, ihr Gesicht zu wahren. Den Stimuluspersonen wurde zu keinem Zeitpunkt angedroht, dass sie die 20€ Belohnung nicht erhalten würden, falls sie den Interviewer nicht überzeugen

könnten; sie nahmen lediglich an, dass dies der Fall sein würde. Somit stand für die Stimulusperson mehr auf dem Spiel als in den meisten Studien (vgl. Vrij, 2008). Dieses Vorgehen wurde aufgrund von ethischen Richtlinien gewählt (vgl. APA Manual, 6. Auflage, in Bezug auf den Schutz von Rechten und Wohlergehen von Forschungsteilnehmern, S. 16 f.). Aus ethischen Gründen ist es schwierig, die Einsätze der Probanden weiter zu erhöhen, da den Teilnehmern zu Beginn jeder Studie mitgeteilt wird, dass sie auf Wunsch die Studie jederzeit abbrechen können (Vrij, Granhag, & Porter, 2010). Während des Interviews stellte der Interviewer allen Stimuluspersonen dieselben Fragen: (a) in Bezug auf ihre Strategien zur Lösung der Aufgaben, (b) in Bezug auf Erklärungen für ihre sehr gute Leistung und (c) ob sie unerlaubte Hilfsmittel verwendet hatten. Die Interviews dauerten im Durchschnitt 2.22 Minuten ($SD = 0.15$, $Min = 1.96$, $Max = 2.43$). Insgesamt wurden 90 Videos produziert²⁴; aus diesen wurden zufällig 24 Aufnahmen ausgewählt. Die ausgewählten Videos wurden drei Sets zugeordnet. Jedes Set beinhaltete acht Videoaufnahmen mit jeweils vier weiblichen und vier männlichen Stimuluspersonen und vier wahren und vier erfundenen Aussagen. Die Sets waren gleich lang und dauerten im Durchschnitt 17.74 Minuten.

²⁴ An dieser Stelle möchte ich Liborio Cicarello, Elina Essig, Julia Dahm, Anwar Helmcke, Max Götz, Marius Marksteiner, Simon von Schwedler, Johannes Hartmann, Saskia Mitreuter, Ann Seibert und Miriam Hippe für ihre tatkräftige Unterstützung und ihr außerordentlich hohes Engagement bei der Erstellung der Videoaufnahmen danken.

5.2.3 *Versuchsdesign*

Experiment 2 folgte wie Experiment 1 einem 2 (Urteilsvermeidung: wahr vs. erfunden) x 2 (Hinweisreiz: wahr vs. erfunden) faktoriellen Design mit den Zwischensubjektfaktoren „Urteilsvermeidung“ und „Hinweisreiz“ und dem Innersubjektfaktor „Valenz der Aussage“.

5.2.4 *Stichprobe*

An Experiment 2 nahmen insgesamt 96 Lehrkräfte teil. Dreiundzwanzig Probanden (23.95%) mussten aufgrund nicht-gegebener Urteilsbegründungen und fehlender zusätzlicher Begründungen aus den Analysen ausgeschlossen werden. Wie bereits erwähnt wurden die Probanden gebeten, jedes Glaubwürdigkeitsurteil direkt im Anschluss an das jeweilige Interview kurz zu begründen (*Urteilsbegründung*) und in Abhängigkeit ihrer Urteilsvermeidung und ihres Urteils eine ausführliche *zusätzliche Begründung* abzugeben. Sofern sie keine Urteilsbegründung ($N = 4$) und/oder zusätzliche Begründung abgaben ($N = 19$), folgten sie nicht den Instruktionen und respektierten somit nicht die Voraussetzungen für die Teilnahme an der Online-Studie. Lediglich eine Person bestand nicht den Instruktionscheck für die Urteilsvermeidung und wurde daher ebenfalls aus den Analysen ausgeschlossen. Mehr als die Hälfte der Teilnehmer war weiblich (63.0%), was unter Lehrern als normale Verteilung angesehen werden kann. Die Teilnehmer arbeiteten entweder an einer weiterführenden Schule (72.6%) oder einer beruflichen Schule (27.4%). Im Durchschnitt waren die Teilnehmer 41.88 Jahre ($SD = 10.95$) alt und im Mittel bereits seit 12.17 Jahren ($SD = 11.04$) im Schuldienst.

5.2.5 *Abhängige Variablen*

Als abhängige Variablen wurden das Glaubwürdigkeitsurteil bzw. Wahrheitsverzerrung, die Akkuratheit bzw. die Trefferquote und die Trefferquote für wahre und erfundene Aussagen erfasst. Die Trefferquote wurde berechnet, indem für jedes Urteil ein Punkt vergeben wurde, wenn Wahrheit oder Lüge korrekt identifiziert werden konnte, und kein Punkt vergeben wurde, wenn die Aussage nicht richtig als wahr oder erfunden eingestuft werden konnte. Somit konnte jeder Teilnehmer zwischen 0 und 8 Punkten erzielen. Die Gesamtpunktzahl jedes Teilnehmers wurde durch die Gesamtzahl aller Urteile (jedes Set beinhaltete acht Aussagen, die beurteilt werden konnten) dividiert und mit 100 multipliziert. Die Trefferquote für wahre Aussagen wurde ähnlich berechnet; hier wurde die Gesamtpunktzahl (die Teilnehmer konnten zwischen 0 und 4 Punkten erreichen) für das korrekte Erkennen von wahren Aussagen durch die Anzahl tatsächlich wahrer Aussagen (in jedem Set waren vier Aussagen wahr) dividiert und mit 100 multipliziert. Analog dazu wurde die Trefferquote für erfundene Aussagen berechnet.

Die Trefferquote für wahre und erfundene Aussagen kann durch eine Wahrheitsverzerrung oder Lügenverzerrung beeinflusst sein, d.h. Urteiler, die dazu tendierten die Aussagen als wahr einzustufen (d.h. Teilnehmer mit einer Wahrheitsverzerrung), haben eine größere Wahrscheinlichkeit, eine höhere Trefferquote für wahre Aussagen und eine geringere Trefferquote für erfundene Aussagen zu erzielen. Analoges gilt für Teilnehmer mit einer Lügenverzerrung. Aufgrund dessen wurde die Valenz des Urteils berücksichtigt, indem die Variable „Wahrheitsverzerrung“ gebildet wurde. Die Urteilsverzerrung wurde

berechnet, indem für jeden Urteiler aufaddiert wurde, wie häufig das Urteil „die Aussagen sind wahr“ abgegeben wurde. Insoweit muss beachtet werden, dass die allgemeine Trefferquote (Trefferquote für wahre plus erfundene Aussagen) nicht von der Urteilsverzerrung beeinflusst wird, da beispielsweise ein Teilnehmer mit einer ausgeprägten Wahrheitsverzerrung (d.h. er stuft alle Aussagen als wahr ein) eine Trefferquote von 100% bei wahren Aussagen, bei erfundenen Aussagen eine Trefferquote von 0% und eine allgemeine Trefferquote von 50% hat. Eine allgemeine Trefferquote von 50% hätte (aufgrund der gleichen Anzahl von wahren und erfundenen Aussagen pro Set) ebenso erreicht werden können, indem bei jedem Urteil geraten wurde, wobei solche Urteiler eine Trefferquote bei wahren und erfundenen Aussagen von jeweils 50% hätten.

5.3 Ergebnisse

5.3.1 Wahrheitsverzerrung

Im Durchschnitt wurden $M = 4.83$ ($SD = 1.42$) Aussagen als wahr eingestuft. Dieser Mittelwert ist signifikant höher als 4, $t(72) = 5.01$, $p < .001$. Um zu testen, ob die Manipulation des Urteilsziels dazu führte, dass die Teilnehmer die Aussagen als tendenziell wahr oder erfunden einstufen und somit die Urteilsverzerrung beeinflusst wurde, wurde eine ANOVA mit „Wahrheitsverzerrung“ als abhängiger Variable und „Urteilsvermeidung“ und „Hinweisreiz“ als unabhängigen Variablen durchgeführt. Die Ergebnisse zeigten einen signifikanten Haupteffekt der Urteilsvermeidung auf die Wahrheitsverzerrung, $F(1, 69) = 4.19$, $p = .045$, $\eta^2 = .057$. Somit stuften die Probanden der Bedingung „Vermeidung Urteil erfunden“ signifikant häufiger

die Aussagen als wahr ein ($M = 5.17$, $SD = 1.50$) und die Probanden der Bedingung „Vermeidung Urteil wahr“ diese häufiger als erfunden ein ($M = 4.51$, $SD = 1.28$). Alle anderen Effekte waren nicht signifikant ($ps > .33$).

5.3.2 Trefferquote

Die durchschnittliche Trefferquote der Probanden lag bei 50.17% ($SD = 13.58$) und unterschied sich nicht signifikant vom Zufallsniveau²⁵ 50%, $t(72) = 0.11$, $p = .92$. Somit waren die Probanden nicht besser als der Zufall im Erkennen von wahren und erfundenen Aussagen. Des Weiteren war die Trefferquote nicht signifikant mit Geschlecht, Alter oder beruflicher Erfahrung (in Jahren) korreliert, alle $ps > .07$. Um den Effekt von Urteilsvermeidung und urteilsziel(in)konsistentem Hinweisreiz zu testen, wurde eine ANOVA mit „Trefferquote“ als abhängiger Variable und „Urteilsvermeidung“ und „Hinweisreiz“ als unabhängigen Variablen durchgeführt. Die Ergebnisse deuten darauf hin, dass der erwartete Interaktionseffekt von Urteilsvermeidung und Hinweisreiz auf die Trefferquote signifikant war, $F(1, 69) = 3.97$, $p = .050$, $\eta^2 = .054$. Keine weiteren Effekte waren signifikant ($ps > .20$). Somit waren die Probanden in der Bedingung „Vermeidung Urteil erfunden“ besser im Unterscheiden von Wahrheit und Lüge, nachdem sie einen urteilszielinkonsistenten Lügenhinweisreiz erhalten hatten ($M = 50.74\%$, $SD = 12.86$) als nach dem Erhalt des urteilszielkonsistenten Wahrheitshinweisreizes ($M = 45.39\%$, $SD = 14.56$). Probanden in der Bedingung „Vermeidung Urteil

²⁵ Wie bereits oben erläutert wird unter Zufallsniveau die Ratewahrscheinlichkeit verstanden, wenn 50% wahre und 50% erfundene Aussagen zur Beurteilung vorgelegt werden (vgl. Vrij, 2008, S. 141).

wahr“ waren ebenfalls akkurater, nachdem sie den urteilszielinkonsistenten (Wahrheits-)Hinweisreiz erhalten hatten ($M = 55.63$, $SD = 13.74$) als nach dem urteilszielkonsistenten (Lügen-)Hinweisreiz ($M = 48.53$, $SD = 11.59$). Simple-Effect-Analysen nach Field (2009) deuten darauf hin, dass bei einem Hinweisreiz „wahr“ Lehrpersonen in der Bedingung „Vermeidung Urteil wahr“ eine signifikant höhere Trefferquote erzielten im Vergleich zu Probanden der Bedingung „Vermeidung Urteil erfunden“, $F(1, 69) = 5.77$, $p = .02$. Bei einem Hinweisreiz „erfunden“ gab es keine signifikanten Unterschiede zwischen den beiden Urteilsvermeidungs-Bedingungen, $F < 1$. Weitere Simple-Effect-Analysen deuten darauf hin, dass es weder in der Bedingung „Vermeidung Urteil erfunden“ noch in der Bedingung „Vermeidung Urteil wahr“ signifikante Unterschiede zwischen der Darbietung eines urteilszielkonsistenten vs. – inkonsistenten Hinweisreizes gibt, alle $ps > .10$. Der Interaktionseffekt ist in Abbildung 5 dargestellt.

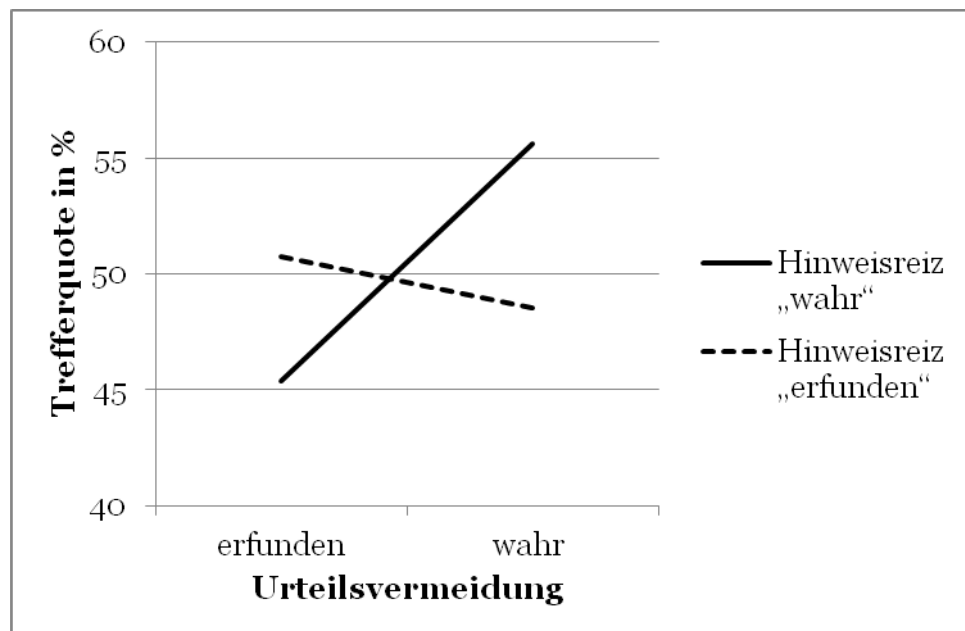


Abbildung 5. Trefferquote in Prozent in Abhängigkeit von der Urteilsvermeidung und des Hinweisreizes in Experiment 2.

5.3.3 *Trefferquote bei wahren und erfundenen Aussagen*

Um zu überprüfen, ob Probanden besser im Erkennen von wahren oder erfundenen Aussagen waren, wurde eine ANOVA mit Messwiederholung durchgeführt mit „Valenz der Aussage“ (Trefferquote bei wahren und erfundenen Aussagen) als Innersubjektfaktor und „Urteilsvermeidung“ und „Hinweisreiz“ als Zwischensubjektfaktoren. Wie erwartet waren die Probanden signifikant besser im Erkennen von wahren Aussagen ($M = 60.62\%$, $SD = 22.41$) als im Erkennen von erfundenen Aussagen ($M = 39.73\%$, $SD = 22.37$), $F(1, 69) = 25.88$, $p < .001$, $\eta^2 = .27$. Beide Mittelwerte unterschieden sich signifikant vom Zufallsniveau 50%, $t_{\text{erfunden}}(72) = -3.92$, $p_{\text{erfunden}} < .001$ und $t_{\text{wahr}}(72) = 4.05$, $p_{\text{wahr}} < .001$. Somit waren sie besser als der Zufall beim Erkennen von wahren Aussagen und signifikant schlechter als der Zufall im Erkennen von erfundenen Aussagen. Dieser Haupteffekt des Innersubjektfaktors wurde – wie erwartet – qualifiziert durch eine signifikante Interaktion zwischen Valenz der Aussage und Urteilsvermeidung, $F(1, 69) = 4.96$, $p = .029$, $\eta^2 = .068$, alle anderen F s < 1 . Simple-Effect-Analysen (vgl. Field, 2009, Syntax siehe Anhang II.a) zeigen, dass Probanden der Bedingung „Vermeidung Urteil erfunden“ nicht signifikant besser im Erkennen von wahren Aussagen waren ($M = 62.50\%$, $SD = 23.53$) als Probanden der Bedingung „Vermeidung Urteil wahr“ ($M = 58.78\%$, $SD = 21.42$), $F < 1$. Dagegen waren Probanden der Bedingung „Vermeidung Urteil wahr“ signifikant besser im Erkennen von erfundenen Aussagen ($M = 45.95\%$, $SD = 19.99$) als Probanden der Bedingung „Vermeidung Urteil erfunden“ ($M = 33.33\%$, $SD = 23.15$), $F(1, 71) = 6.22$, $p = .015$, $\eta^2 = .081$.

5.4 Diskussion Experiment 2

In Experiment 2 lag der Fokus auf dem Effekt des Urteilsziels und des Verarbeitungsmodus auf die Fähigkeit, akkurat wahre und erfundene Aussagen zu unterscheiden. Es wurde angenommen, dass Probanden der Bedingung „Vermeidung Urteil erfunden“ mehr Aussagen als wahr einstufen und Probanden der Bedingung „Vermeidung Urteil wahr“ mehr Aussagen als erfunden einstufen (vgl. C. F. Bond & DePaulo, 2006; Reinhard, Dickhäuser, et al., 2011). Die Ergebnisse stützen diese Annahme. Als ein Ergebnis der Verzerrung wurde davon ausgegangen, dass Probanden der Bedingung „Vermeidung Urteil erfunden“ wahre Aussagen besser als solche erkennen und Probanden der Bedingung „Vermeidung Urteil wahr“ erfundene Aussagen besser als solche erkennen (vgl. C. F. Bond & DePaulo, 2006; Reinhard, Dickhäuser, et al., 2011). Die Ergebnisse deuten darauf hin, dass Probanden beider Bedingungen Aussagen eher als richtig einstufen, wenn diese ihrer Urteilsvermeidung entsprechen. Des Weiteren wurde angenommen, dass Probanden in beiden Bedingungen beim Erkennen von wahren und erfundenen Aussagen auf Zufallsniveau sind (vgl. C. F. Bond & DePaulo, 2006; Reinhard, Dickhäuser, et al., 2011). Diese Annahme konnte ebenfalls durch die Ergebnisse bestätigt werden. Die Hauptannahme – dass Probanden nach Erhalt eines urteilszielinkonsistenten Hinweisreizes besser erfundene von wahren Aussagen unterscheiden können – konnte ebenfalls bestätigt werden. Dies kann dadurch erklärt werden, dass eine systematische Verarbeitung zu einer Nutzung verbaler Informationen bzw. des Nachrichteninhalts führt (Reinhard & Sporer, 2008, 2010) und mehr verbale Hinweisreize als valide Lügenindikatoren gelten

(DePaulo et al., 2003). Wichtig ist hierbei, dass dieser Effekt unabhängig vom Urteilsziel auftrat; dies kann dadurch begründet werden, dass eine verzerrte systematische Verarbeitung eine selektive Verarbeitung von eher validen, verbalen Informationen nach sich zieht (Chen & Chaiken, 1999).

6 Mehrebenenanalysen und Glaubwürdigkeitsforschung

In Experiment 2 wurde der Einfluss von Urteilsziel und Verarbeitungsmodus auf die Urteilsverzerrung (Wahrheitsverzerrung) und die Akkuratheit beim Erkennen von Lüge und Wahrheit (Trefferquote) untersucht. Dabei zeigte sich, dass das Urteilsziel sich auf die Urteilsverzerrung auswirkt und eine „Vermeidung Urteil erfunden“ zu einer stärkeren Wahrheitsverzerrung führte als eine „Vermeidung Urteil wahr“. Des Weiteren zeigte sich, dass ein urteilszielinkonsistenter Hinweisreiz zu einer höheren Trefferquote führt als ein urteilszielkonsistenter Hinweisreiz. Bei beiden abhängigen Variablen, sowohl bei der Wahrheitsverzerrung als auch bei der Trefferquote, handelte es sich um Summenwerte, d.h. um aggregierte Daten. Da Aggregation von Daten zu Informationsverlust führt (vgl. Bortz & Döring, 2006; Twisk, 2011), würde eine Messwiederholungsanalyse ein präziseres statistisches Urteil erlauben. Zudem ist die Nichtberücksichtigung von Einflüssen von Schülervariablen in Experiment 2 als problematisch anzusehen, da neben Merkmalen des Senders (z.B. Urteilsziel und Verarbeitungsmodus) auch Merkmale der Quelle (z.B. glaubwürdiges Auftreten des Senders) das Glaubwürdigkeitsurteil beeinflussen können (vgl. sender demeanor, Levine et al., 2011; Vrij, 2008). Eine Möglichkeit, diesen Problemen entgegenzuwirken, ist eine Modellierung der Daten anhand eines (logistischen) Mehrebenenregressionsmodells. Eine Mehrebenenregression ist einer Varianzanalyse (ANOVA) überlegen: Bei einer ANOVA mit Messwiederholung können im Vergleich zur Mehrebenenregression (1) keine Variablen erster Ebene (z.B. Schülermerkmale) und zweiter Ebene (z.B. Lehrermerkmale) gleichzeitig berücksichtigt werden (ohne Variablen

experimentell variiert zu haben) und (2) der Einfluss metrischer Prädiktoren kann ohne Transformation der Daten (z.B. Median-Split) nicht überprüft werden, was zu Informationsverlust führt.

Aufgrund der genannten Vorteile von Mehrebenenanalysen (für eine ausführliche Erläuterung siehe z.B. Twisk, 2011) wurde eine solche in Experiment 3 durchgeführt. Zum besseren Verständnis werden im Folgenden zunächst die Grundlagen der Mehrebenenanalyse dargestellt, bevor anschließend auf die Möglichkeit der Nutzung dieses Analyseverfahrens in der Glaubwürdigkeitsforschung eingegangen wird.

6.1 Grundlagen der Mehrebenenanalyse

6.1.1 Genestete und nicht-genestete Modelle

Genestete Modelle. Klassischerweise werden Mehrebenenanalysen mit Datensätzen durchgeführt, bei denen eine hierarchische Struktur existiert: Es gibt mehrere Ebenen und bestimmte Einheiten sind in höheren Einheiten genestet (vgl. Hox, 2010; Twisk, 2011; siehe Abbildung 6). So können z.B. Schüler (Einheit) in Klassen (höhere Einheit) geschachtelt sein. Auf Ebene 1 eines solchen Modells wären die Schüler, auf Ebene 2 die Klassen angesiedelt. Dabei wird angenommen, dass die Einheiten einer Ebene in gewisser Weise voneinander abhängig sind. So wird z.B. davon ausgegangen, dass Schüler einer Klasse sich hinsichtlich bestimmter Merkmale ähnlicher sind als Schülern einer anderen Klasse. Zum einen kann diese Annahme durch Befunde aus dem Bereich des sozialen Einflusses durch Gruppen unterstützt werden (vgl. Aronson et al., 2008), z.B. werden Schüler durch Mitschüler, die sich in der gleichen Klasse befinden, die sie täglich sehen und mit denen sie auf

Schulausflüge gehen, stärker beeinflusst als durch Mitschüler, die sich in einer anderen Klasse befinden und mit denen sie weniger Zeit gemeinsam verbringen. Zum anderen kann diese Merkmalsähnlichkeit durch Selektion entstanden sein, z.B. indem Schüler, die sich eher für naturwissenschaftliche Fächer interessieren, in einer und Schüler, die sich eher für Sprachen interessieren, in einer anderen Klasse sind.

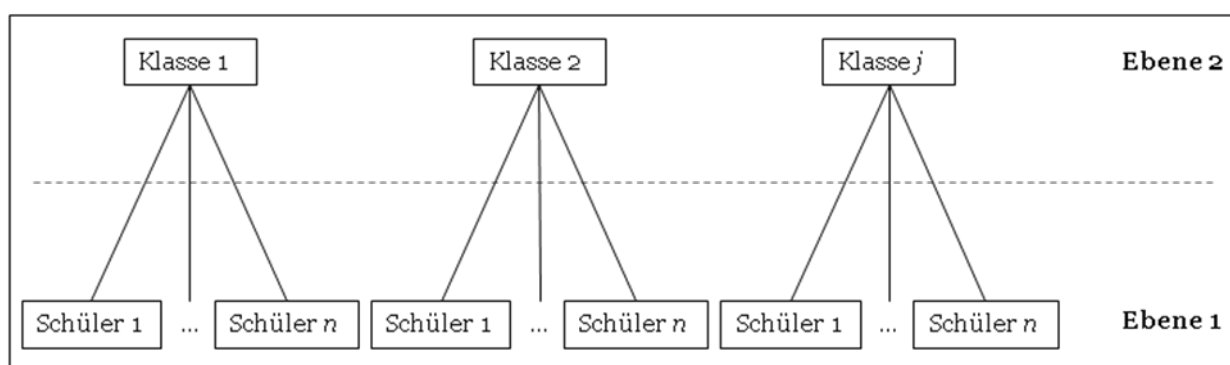


Abbildung 6. Klassische Struktur eines genesteten Modells mit zwei Ebenen: Auf Individualebene befinden sich die Schüler (Ebene 1), die in Klassen (Ebene 2) geschachtelt sind.

Nicht-genestete Modelle. Neben diesen hierarchischen bzw. genesteten Datensätzen gibt es auch nicht-hierarchische bzw. nicht-genestete Datensätze (vgl. Gelman & Hill, 2009). Dies ist z.B. dann der Fall, wenn ein Lehrer mehrere Schüler unterrichtet und beurteilt, diese Schüler aber auch von anderen Lehrern unterrichtet und beurteilt werden (vgl. Abbildung 7). Bei dieser Messwiederholung (aufgrund der wiederholten Beurteilung) befinden sich die Schüler (bzw. die Urteile über die Schüler) auf Ebene 1 und die Lehrer auf Ebene 2 des Modells (vgl. Abbildung 7). Eine Messwiederholung kann also in ein nicht-genestetes Mehrebenenmodell überführt werden, bei dem Urteile (z.B. über Schüler) in Untersuchungseinheiten (z.B. Lehrer) geschachtelt sind (Gelman & Hill, 2009). Dadurch, dass ein Lehrer mehrere Schüler beurteilt, die gleichzeitig von anderen Lehrern beurteilt werden, ist eine eindeutige

Zuordnung von Einheiten erster Ebene zu einer einzelnen Einheit zweiter Ebene allerdings nicht mehr möglich ist (z.B. kann ein Schüler nicht mehr eindeutig einem einzigen Lehrer, sondern mehreren Lehrern zugeordnet werden). Aufgrund dieser Überschneidungen hinsichtlich der Zuordnung spricht man von einem nicht-genesteten Modell.

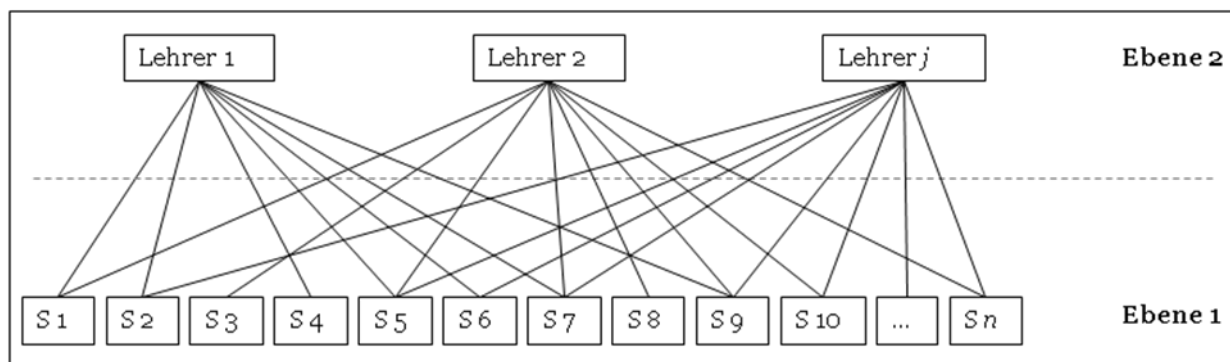


Abbildung 7. Struktur eines nicht-genesteten Modells mit zwei Ebenen: Auf Individualebene befinden sich die Urteile über die Schüler (Ebene 1), die mehreren Lehrern (Ebene 2) zugeordnet sind.

6.1.2 Lineare, logistische und Mehrebenen-Regression

Einfache lineare Regression und lineare Mehrebenenregression. Die Mehrebenenregression basiert auf der *einfachen linearen Regression*, bei der ein Kriterium (z.B. das auf einer Skala von 1 bis 10 erfasste Glaubwürdigkeitsurteil über einen Schüler) durch einen oder mehrere Prädiktoren (z.B. die Länge der Antwort des Schülers) vorhergesagt wird. Wie bereits erwähnt existieren bei der *Mehrebenenregression* mehrere Ebenen, wobei die Einheiten der ersten Ebene (Schülerebene) Einheiten einer höheren Ebene (Lehrerebene) zugeordnet werden (z.B. können Urteile über Schüler bestimmten Lehrern zugeordnet werden). Wie bei der einfachen linearen Regression werden bei der Mehrebenenregression auf der ersten Ebene Zusammenhänge zwischen einem Prädiktor und einem Kriterium vorhergesagt

(z.B. der Zusammenhang zwischen dem Glaubwürdigkeitsurteil über einen Schüler und dessen Antwortlänge). Diese Zusammenhänge können durch Prädiktoren zweiter Ebene moderiert werden (z.B. durch das Urteilsziel und den Verarbeitungsmodus des Lehrers). Somit beeinflussen die Prädiktoren der zweiten Ebene, wie stark der Zusammenhang zwischen dem Kriterium und dem Prädiktor erster Ebene ist. Beispielsweise kann der Zusammenhang zwischen dem Glaubwürdigkeitsurteil über einen Schüler (Kriterium) und dessen Antwortlänge (Prädiktor erster Ebene) stärker bei Lehrern sein, welche die Schülerinformationen systematisch (vs. heuristisch) verarbeiten (Prädiktor zweiter Ebene). Da der Zusammenhang zwischen Prädiktor und Kriterium der ersten Ebene unterschiedlich stark für die verschiedenen Untersuchungseinheiten zweiter Ebene sein können (aufgrund von Prädiktoren der zweiten Ebene), wird bei der Mehrebenenregression für jede Einheit der zweiten Ebene eine separate Regressionsgleichung aufgestellt. Beispielsweise wird für jeden Lehrer eine Regressionsgleichung aufgestellt, die den Zusammenhang zwischen dem Glaubwürdigkeitsurteil und der Antwortlänge veranschaulicht (Hox, 2010; Twisk, 2011).²⁶

Neben dieser Annahme hinsichtlich der unterschiedlichen Stärke der Zusammenhänge kann in einer Mehrebenenregression auch berücksichtigt werden, dass die durchschnittliche Ausprägung des Kriteriums für die

²⁶ Wird von Unterschieden zwischen den Untersuchungseinheiten zweiter Ebene hinsichtlich der Steigung (Slopes) der Regressionsgerade bzw. des Zusammenhangs zwischen Prädiktor und Kriterium erster Ebene ausgegangen, wird von einem „Random Slopes“-Modell gesprochen (Hox, 2010; Twisk, 2011).

Untersuchungseinheiten der zweiten Ebene unterschiedlich groß ist (wenn der Prädiktor erster Ebene den Wert „0“ annimmt). Somit kann berücksichtigt werden, wenn das Intercept (Y-Achsen-Abschnitt) zwischen den Einheiten zweiter Ebene unterschiedlich ausgeprägt ist. Beispielsweise kann angenommen werden, dass das durchschnittliche Glaubwürdigkeitsurteil (wenn die Antwortlänge den Wert „0“ angenommen hat) bei Lehrer A höher ist als bei Lehrer B und somit das Intercept für Lehrer A höher ist als für Lehrer B (Hox, 2010; Twisk, 2011).²⁷

Logistische Regression und logistische Mehrebenenregression. Bisher wurde lediglich der Fall metrischer Kriteriumsvariablen betrachtet. In vielen Fällen will man aber keine metrische Variable (wie die Glaubwürdigkeit eines Schülers gemessen auf einer Skala von 1 bis 10 oder dessen schulische Leistung) vorhersagen, sondern Effekte auf eine kategoriale bzw. dichotome Kriteriumsvariable schätzen. Z.B. soll der Einfluss eines Prädiktors auf das Urteil getestet werden, ob ein Schüler lügt oder die Wahrheit sagt oder ob er in eine höhere Klassenstufe versetzt wird oder nicht. Um dichotome Kriteriumsvariablen in Mehrebenenmodellen vorherzusagen, werden logistische Regressionsanalysen durchgeführt. Im Unterschied zur linearen Regression und linearen Mehrebenenregression wird bei einer logistischen Regression und einer logistischen Mehrebenenregression die Wahrscheinlichkeit für ein bestimmtes Ereignis vorhergesagt. So wird z.B. die Wahrscheinlichkeit, die Aussagen eines Schülers als „wahr“ im Vergleich zu „erfunden“ zu beurteilen,

²⁷ Geht man von einer unterschiedlichen Ausprägung des Intercepts aus, spricht man von einem „Random-Intercept“-Modell (Hox, 2010; Twisk, 2011).

oder die Wahrscheinlichkeit, ein richtiges im Vergleich zu einem falschen Urteil zu treffen, vorhergesagt. Bei einer logistischen (Mehrebenen-)Regression ist die Kriteriumsvariable das *Logit der Wahrscheinlichkeit für ein bestimmtes Ereignis*, z.B. das Logit der Wahrscheinlichkeit eines Lehrers, die Aussagen eines Schülers als „wahr“ zu beurteilen. Ein *Logit* stellt den (natürlichen) Logarithmus²⁸ der Odds dar. *Odds* werden berechnet, indem die Wahrscheinlichkeit für das Eintreten des Ereignisses A durch die Gegenwahrscheinlichkeit bzw. durch die Wahrscheinlichkeit des Eintretens des Ereignisses B dividiert wird. Beispielsweise wird zur Berechnung der Odds die Wahrscheinlichkeit, das Urteil „wahr“ zu geben, ins Verhältnis zur (Gegen-)Wahrscheinlichkeit, das Urteil „erfunden“ zu geben, gesetzt (Hox, 2010; Twisk, 2011).

Zusammenfassung. Die Mehrebenenregression basiert also auf der einfachen Regression, mit dem Unterschied, dass bei der Mehrebenenregression für jede Einheit der zweiten Ebene eine Regressionsgleichung aufgestellt werden kann und dadurch Einflüsse der Prädiktoren zweiter Ebene auf den Zusammenhang zwischen Prädiktor und Kriterium erster Ebene berücksichtigt werden können. Neben der Moderation dieser Zusammenhänge kann bei einer Mehrebenenregression auch berücksichtigt werden, dass die Einheiten zweiter Ebene – unabhängig vom

²⁸ Ein *Logarithmus* (z.B. $\log_2 8 = 3$ [Logarithmus von 8 zur Basis 2 ist gleich 3]) ist die Umkehroperation des Potenzierens: Beim *Potenzieren* wird ein Faktor wiederholt mit sich selbst multipliziert (z.B. $2^3 = 2 \cdot 2 \cdot 2 = 8$). Der *natürliche* Logarithmus ist der Logarithmus zur Basis e (Eulersche Zahl = 2.718; Bortz & Döring, 2006).

Prädiktor erster Ebene – eine unterschiedliche durchschnittliche Ausprägung des Kriteriums haben können. Des Weiteren können wie bei einfachen Regressionsanalysen auch bei Mehrebenenregressionsanalysen Einflüsse von Prädiktoren auf metrische und auf kategoriale Kriteriumsvariablen untersucht werden.

6.2 Mehrebenenanalysen in der Glaubwürdigkeitsforschung

Einflüsse des Senders auf das Glaubwürdigkeitsurteil. Bisherige Studien konnten zeigen, dass die Glaubwürdigkeitsbeurteilung nicht nur durch Merkmale des Empfängers beeinflusst wird, z.B. durch die Wahrheitsverzerrung (vgl. C. F. Bond & DePaulo, 2006), sondern auch durch Merkmale des Senders, z.B. durch die logische Konsistenz einer Aussage (Reinhard & Sporer, 2008, 2010). Eine aktuelle Metaanalyse von Hartwig und Bond (2011) deutet darauf hin, dass Personen sich bei der Glaubwürdigkeitsbeurteilung auf Merkmale des Senders stützen, z.B. die Redezeit (*talking time*) oder die logische Struktur (*logical structure*). Daher scheint es sinnvoll, neben Eigenschaften des Empfängers bzw. Urteilers auch Eigenschaften des Senders bzw. des Beurteilten bei den Analysen zu berücksichtigen. Eine Möglichkeit, Eigenschaften von Urteiler und Beurteiltem gleichzeitig zu berücksichtigen, ist die Analyse der Daten anhand einer Mehrebenenanalyse.

Aggregation abhängiger Variablen in der Glaubwürdigkeitsforschung. In bisherigen Studien zur Untersuchung des Erkennens von Lüge und Wahrheit (vgl. C. F. Bond & DePaulo, 2006) werden Probanden klassischerweise ca. 16 auf Video aufgenommene Aussagen von Personen zur Beurteilung vorgelegt, wobei jede Aufnahme durchschnittlich 52 Sekunden lang ist. Häufig werden die

Videoaufnahmen auch in verschiedene Sets eingeteilt (z.B. Marksteiner, Reinhard, et al., 2012; Reinhard, Dickhäuser, et al., 2011) und jeder Proband erhält per Zufall ein Set zur Beurteilung. Auf diese Weise können insgesamt mehr Aussagen beurteilt werden, ohne dass ein Proband alle Aufnahmen beurteilen muss. Um die Einflüsse auf die Akkuratheit eines Glaubwürdigkeitsurteils und die Urteilsverzerrung testen zu können, werden meistens die Trefferquote und die Wahrheitsverzerrung durch einen Summenscore berechnet (siehe Experiment 2 zur Berechnung der Trefferquote und der Urteilsverzerrung). Der Nachteil an dieser Vorgehensweise ist, dass durch das Aufsummieren bzw. Aggregieren der Daten einer Messwiederholung wertvolle Informationen verloren gehen (vgl. Bortz & Döring, 2006; Twisk, 2011).

Um eine Aggregation der Daten zu vermeiden, kann eine Messwiederholung in ein nicht-genestetes Mehrebenenmodell überführt werden, bei dem Urteile in Versuchspersonen geschachtelt sind (vgl. Gelman & Hill, 2009). Da in den meisten Experimenten zur Glaubwürdigkeitsforschung mehrere auf Video aufgenommene Aussagen zur Beurteilung vorgelegt werden, handelt es sich häufig um eine Messwiederholung, deren Daten sinnvoll in einer Mehrebenenregressionanalyse untersucht werden können.

Skalenniveau abhängiger Variablen in der Glaubwürdigkeitsforschung.

Um zu erforschen, ob Personen akkurat zwischen wahren und erfundenen Aussagen unterscheiden können, werden die Urteile über die Glaubwürdigkeit einer Person klassischerweise binär erfasst (vgl. Bond & DePaulo, 2006). Zur Analyse von Einflüssen auf kategoriale abhängige Variablen, z.B. auf die

Akkuratheit eines Glaubwürdigkeitsurteils („Urteil richtig/falsch“), werden logistische Mehrebenenregression durchgeführt.

Zusammenfassung. Aufgrund des gleichzeitigen Einflusses des Urteilers und des Beurteilten auf das Glaubwürdigkeitsurteil, der Struktur der Daten aus der Glaubwürdigkeitsforschung (Messwiederholung) und aufgrund des Skalenniveaus der abhängigen Variablen („Urteil wahr/erfunden“ bzw. „Urteil richtig/falsch“) scheint es sinnvoll, diese anhand einer logistischen Mehrebenenregression auszuwerten.

7 Experiment 3: Einfluss von unbewussten Urteilszielen und Verarbeitungsmodus auf die Akkuratheit von Glaubwürdigkeitsurteilen unter der Berücksichtigung von Schülermerkmalen

7.1 Fragestellung und Hypothesen

In Experiment 2 wurde untersucht, inwieweit sich Urteilsziel und Verarbeitungsmodus auf die Akkuratheit von Glaubwürdigkeitsurteilen auswirken. Experiment 3 dient neben der Replikation der Befunde aus Experiment 2 der Untersuchung des Einflusses von Schülermerkmalen auf die Akkuratheit von Glaubwürdigkeitsurteilen. Zur Testung dieses Einflusses werden die Daten mehrebenenanalytisch ausgewertet. Durch die Analyse der Daten anhand einer logistischen Mehrebenenregressionsanalyse wird zum einen dem Informationsverlust, der durch Aggregation der Daten zustande kommt, entgegengewirkt. Des Weiteren können durch eine mehrebenenanalytische Auswertung Variablen auf höherer Ebene berücksichtigt werden (vgl. Kapitel 6).

Wie auch in den vorangegangenen Experimenten wird in Experiment 3 angenommen, dass das Urteilsziel sich auf das Glaubwürdigkeitsurteil auswirkt. Wie auch in Experiment 2 wird im vorliegenden Experiment das Glaubwürdigkeitsurteil anhand eines binären Items (1 = „Die Aussagen sind wahr.“, 2 = „Die Aussagen sind erfunden.“) erfasst. Es wird davon ausgegangen, dass die Probanden in der Bedingung „Vermeidung Urteil erfunden“ mit größerer Wahrscheinlichkeit die Aussagen eines Schülers als wahr einstufen als Probanden der Bedingung „Vermeidung Urteil wahr“ (Hypothese 1). Weiterhin

Experiment 3: Einfluss von unbewussten Urteilszielen und Verarbeitungsmodus auf die Akkuratheit von Glaubwürdigkeitsurteilen unter der Berücksichtigung von Schülermerkmalen

wird wie auch in Experiment 2 davon ausgegangen, dass eine systematische Verarbeitung – induziert durch einen urteilszielinkonsistenten Hinweisreiz (vgl. Giner-Sorolla & Chaiken, 1997) – zu einer höheren Trefferquote führt als eine heuristische Verarbeitung (Masip et al., 2009; Reinhard, Sporer, et al., 2011; Hypothese 2a und 2b für die beiden Vermeidungs-Bedingungen). Dies wird dadurch erklärt, dass eine systematische Verarbeitung zu einer Nutzung verbaler Informationen bzw. Merkmale des Nachrichteninhalts führt (Reinhard & Sporer, 2008, 2010) und mehr verbale Hinweisreize als valide Lügenindikatoren gelten (DePaulo et al., 2003). Wichtig ist hierbei, dass eine systematische Verarbeitung unabhängig von der Manipulation des Urteilsziels zu einer höheren Akkuratheit führen sollte, da eine verzerrte systematische Verarbeitung zwar zu einer selektiven Verarbeitung von verbalen Informationen führt (Chen & Chaiken, 1999), diese aber dennoch zu einem größeren Teil valide sind als nonverbale Informationen (DePaulo et al., 2003).

Bisherige Studien konnten zeigen, dass die Beurteilung der Glaubwürdigkeit nicht nur durch Merkmale des Empfängers beeinflusst wird, z.B. durch das Kognitionsbedürfnis (Reinhard, 2010) oder die Aufgabeneinbindung (Reinhard & Sporer, 2008, 2010) oder wie in Experiment 2 und 3 durch Urteilsziel und Verarbeitungsmodus, sondern auch durch Merkmale der Quelle, z.B. glaubwürdiges Auftreten des Senders (vgl. sender demeanor, Levine et al., 2011; Vrij, 2008). Häufig wird argumentiert, dass die Trefferquote beim Erkennen von Lüge und Wahrheit auf dem Zufallsniveau liegt, da die Empfänger der Nachricht Informationen zur Beurteilung heranziehen, die invalide sind (vgl. DePaulo et al., 2003) und somit keine

Experiment 3: Einfluss von unbewussten Urteilszielen und Verarbeitungsmodus auf die Akkuratheit von Glaubwürdigkeitsurteilen unter der Berücksichtigung von Schülermerkmalen

tatsächlichen Indikatoren für eine wahre oder erfundene Nachricht sind. Die Befunde einer aktuellen Meta-Analyse von Hartwig und Bond (2011) deuten jedoch darauf hin, dass Personen sich bei der Glaubwürdigkeitsbeurteilung durchaus auf valide Informationen stützen, z.B. die logische Struktur (logical structure) einer Aussage. Durch die logistische Mehrebenenregressionsanalyse ist es möglich, solche Merkmale der Quelle ebenfalls zu berücksichtigen und sog. Cross-Level-Interaktionen zu testen, d.h. Interaktionen zwischen Variablen erster und zweiter Ebene (z.B. die Interaktion zwischen Verarbeitungsmodus und Antwortlänge der Schüleraussage). Aufgrund bisheriger Befunde (vgl. Giner-Sorolila & Chaiken, 1997; Masip et al., 2009; Reinhard, Sporer, et al., 2011) und der Befunde aus Experiment 1 und 2 kann angenommen werden, dass urteilszielinkonsistente Hinweisreize zu einer systematischen Verarbeitung von Informationen und damit zur Nutzung von verbalen Informationen führen sollen. Daher kann davon ausgegangen werden, dass Probanden in einer Bedingung mit urteilszielinkonsistentem Hinweisreiz bei ihrem Urteil durch verbale Schülervariablen wie z.B. die Antwortlänge beeinflusst werden (Hypothese 3). Insbesondere wird erwartet, dass Personen mit urteilszielinkonsistentem Hinweisreiz die Aussagen des zu beurteilenden Schülers mit größerer Wahrscheinlichkeit als „wahr“ beurteilen (im Vergleich zu „erfunden“), je länger die Antwort ist (vgl. Bond, 2006; Hartwig & Bond, 2011). Es wird also eine Dreifachinteraktion zwischen den Faktoren „Urteilsvermeidung“, „Hinweisreiz“ und „verbaler Schülervariable“ (z.B. Antwortlänge) auf das Urteil erwartet.

7.2 Methode

7.2.1 Versuchsablauf

Der Versuchsablauf war dem in Experiment 2 ähnlich. Die Erhebung der Daten erfolgte computerbasiert (Unipark, www.unipark.info). Den Probanden wurde per E-Mail ein Link zur Studie zugeschickt. Zunächst erhielten die Probanden allgemeine Informationen über die Dauer der Studie, das Ziel der Untersuchung und die Aufwandsentschädigung. Außerdem wurde ihnen Anonymität und die ausschließliche Verwendung der Daten für wissenschaftliche Zwecke zugesichert. Anschließend schauten die Probanden eine kurze Videosequenz, in der ein Mitarbeiter des Lehrstuhls zu sehen war, der die Probanden bat, die Audioeinstellungen an ihrem Computer zu überprüfen.²⁹ Danach wurden die Probanden gebeten Angaben bezüglich ihres Alters, Geschlechts, Berufs und ihrer Berufserfahrung (in Jahren und Monaten) zu machen. Das anschließend präsentierte Szenario war dasselbe wie in Experiment 2. Nach der Präsentation des Szenarios erfolgte die Manipulation des Urteilsziels (Urteilsvermeidung erfunden vs. Urteilsvermeidung wahr) wie in Experiment 2. Des Weiteren wurden die Probanden gebeten, ihr Urteil kurz mit Hilfe weniger Stichworte zu begründen. Die Probanden wurden ebenfalls gebeten, nach Beurteilung aller acht Interviews eine detailliertere, zusätzliche Begründung für alle diejenigen Schüler abzugeben, deren Aussagen die Probanden als erfunden bzw. wahr eingestuft hatten. Wie zuvor wurden die

²⁹ An dieser Stelle möchte ich Florian Lettau für die Erlaubnis zur Benutzung der Videosequenz danken.

Instruktionen anhand einer Graphik dargestellt, um die Instruktionen für die Probanden klar zu gestalten. Danach folgte wie in Experiment 2 der Instruktionscheck für die Urteilszielmanipulation. Allerdings wurde den Probanden, sofern sie eine nicht zutreffende Antwort beim Instruktionscheck machten, die Instruktion bezüglich des Urteilsziels nochmals vorgelegt und es erfolgte ein zweiter Instruktionscheck. Schließlich wurde wie in Experiment 2 der entweder urteilszielkonsistente oder urteilszielinkonsistente Hinweisreiz präsentiert. Anschließend folgten wie in Experiment 2 acht Interviews, allerdings waren andere Schüler in den Videoaufnahmen zu sehen als in Experiment 2. Danach beurteilten die Probanden die Glaubwürdigkeit der Schüleraussagen und begründeten die Aussagen kurz (*kurze Urteilsbegründung*). Wie zuvor wurden die Probanden auch diesmal gebeten, ihre Argumente zu nummerieren. Nach Beurteilung aller acht Schülerinterviews wurden die Probanden gebeten, die detailliertere Begründung (*zusätzliche Begründung*) abzugeben, sofern sie ein entsprechendes Urteil abgegeben hatten. Am Ende wurde den Probanden für ihre Teilnahme gedankt und ihnen wurde die Möglichkeit gegeben, an einer Verlosung von Gutscheinen eines namhaften Internetshops teilzunehmen.

7.2.2 Versuchsmaterial

Insgesamt wurden den Probanden wie im zweiten Experiment Videoaufnahmen von 24 Stimuluspersonen vorgelegt; bei den Videoaufnahmen handelte es sich um 24 der insgesamt 90 produzierten Aufnahmen (siehe Experiment 2 für eine ausführliche Erläuterung der Erstellung des Versuchsmaterials). Aus den insgesamt 90 Aufnahmen wurden zufällig 24

Aufnahmen ausgewählt, die nicht bereits in Experiment 2 verwendet worden waren. Zwölf Stimuluspersonen waren männlich, zwölf waren weiblich. Die Interviews dauerten im Durchschnitt 2.10 Minuten ($SD = 0.64$, $Min = 1.2$, $Max = 4.08$). Die 24 ausgewählten Videos wurden insgesamt sechs Sets zugeordnet. Jedes Set beinhaltete acht Videoaufnahmen mit jeweils vier weiblichen und vier männlichen Stimuluspersonen und vier wahren und vier erfundenen Aussagen, wobei die Videos in den sechs Sets teilweise identisch waren.³⁰ Aufgrund dieser Überschneidung der Sets kann von einem nicht-genesteten Modell gesprochen werden (Hox, 2010). Die Sets waren gleich lang und dauerten 16.83 ($SD = 0.59$) Minuten im Durchschnitt.

Alle 24 Videoaufnahmen wurden hinsichtlich verschiedener Lügenindikatoren (siehe Tabelle 1) von zwei unabhängigen Ratern eingeschätzt. Insgesamt wurden folgende acht Indikatoren zur Einschätzung ausgewählt: Antwortlänge (*response length*; Cohens $\kappa = .35$, $p = .002$), Details (*details*; Cohens $\kappa = 1.00$, $p = .002$), Plausibilität (*plausibility*; Cohens $\kappa = .77$, $p = .001$), logische Struktur (*logical structure*; Cohens $\kappa = 1.00$, $p < .001$),

³⁰ Um alle 24 Videos auf einer gemeinsamen Skala abbilden zu können, war eine Überschneidung der Sets notwendig (sog. Booklet-Design). Wären wie üblich (vgl. Bond & DePaulo, 2006) nur 3 Sets mit jeweils acht unterschiedlichen Videoaufnahmen verwendet worden, dann stünde für jedes Set nur ein Drittel der Stichprobe als Informationsquelle zur Verfügung. Durch die Überschneidung kann man aufgrund der Eigenschaften eines Probanden auch Aussagen über Items (z.B. Schüleraussagen) machen, die dieser Proband nicht bearbeitet bzw. beurteilt hat. Umgekehrt stehen für die Schätzung der Itemparameter alle Probanden zur Verfügung, wodurch eine höhere Power erzielt werden kann (vgl. OECD 2009, S. 90 f., zu Optimal Conditions for Linking Items).

Tabelle 1

Mittlere Ausprägung der für Experiment 3 ausgewählten Lügenindikatoren für wahre (M_{wahr}) sowie erfundene Aussagen (M_{erfunden}) und in Hartwig und Bond (2011) berichtete Zusammenhänge zwischen Indikator und wahrgenommener (r_{Per}) sowie tatsächlicher Täuschung (r_{Dec})

		(Hartwig & Bond, 2011)		Experiment 3			
Name				M_{wahr}	M_{erfunden}		
ID	Lügenindikator	r_{Per}	r_{Dec}	(SD)	(SD)	$F(df)$	p
1	Antwortlänge	-.12*	-.04	237.83 (102.96)	134.92 (54.22)	9.39 (22)	.006
4	Details	-.37**	-.20	5.00 (2.56)	3.42 (2.02)	2.83 (22)	.11
12	Plausibilität	-.47**	-.11	4.42 (1.38)	3.67 (1.67)	1.44 (22)	.24
13	Logische Struktur	-.34**	-.16	4.42 (1.24)	3.42 (1.31)	3.68 (22)	.07
28	Blickabwendung	.28**	.05	9.33 (4.68)	7.92 (2.39)	0.87 (22)	.36
61	Nervosität	.30**	.12	4.17 (1.64)	4.08 (1.83)	0.01 (22)	.91

Anmerkung. ID bezieht sich auf die Identifikationsnummer im Anhang A im Artikel von DePaulo et al. (2003). $r_{\text{Per}} = r$ in Bezug auf das mittlere Fisher's Zr für die Assoziation zwischen wahrgenommener Täuschung und Indikator; $r_{\text{Dec}} = r$ in Bezug auf das mittlere Fisher's Zr für die Assoziation zwischen tatsächlicher Täuschung und Indikator. Positive Werte implizieren, dass der Indikator stärker mit (wahrgenommener) Täuschung assoziiert wird (Hartwig & Bond, 2011). M_{wahr} = mittlere Ausprägung des Indikators bei wahren Aussagen in Experiment 3; M_{erfunden} = mittlere Ausprägung des Indikators bei erfundenen Aussagen in Experiment 3. $F(df)$ und p in Bezug auf Experiment 3.

* $p < .05$. ** $p < .001$.

Blickabwendung (*gaze aversion*; Cohens $\kappa = .72$, $p = .004$) und Nervosität (*nervous*; Cohens $\kappa = .74$, $p = .001$). Auf Grundlage der transkribierten Schüleraussagen wurde die Antwortlänge durch die Anzahl der Wörter, welche die Antworten des Senders insgesamt umfasst, eingestuft. Ebenso wurde die Anzahl der Details auf Grundlage der Transkripte ausgezählt. Die Hinweisreize „Plausibilität“, „logische Struktur“ und „Nervosität“ wurden jeweils auf einer 7-stufigen-Likert-Skala mit den Endpunkten 1 (unplausibel/unlogisch/ nicht nervös) und 7 (plausibel/ logisch/ nervös) erfasst. Der Hinweisreiz „Blickabwendung“ wurde erfasst, indem alle zehn Sekunden notiert wurde, ob der Sender den Blick vom Interviewer abwendete oder nicht. Die Blickabwendungen pro zehn Sekunden wurden anschließend aufaddiert. Eine ausführliche Erläuterung sowie die Art der Erfassung befinden sich in Anhang III.a.

Die genannten Indikatoren wurden ausgewählt, da sie alle laut der Metaanalyse von Hartwig und Bond (2011) von Urteilern bei der Glaubwürdigkeitsbeurteilung tatsächlich als Lügenindikatoren genutzt werden (vgl. Tabelle 1 r_{Per}). Des Weiteren zeigt sich, dass die Hinweisreize „Details“, „Plausibilität“, „logische Struktur“ und „Nervosität“ „richtig“ genutzt werden (Hartwig & Bond, 2011): So wird eine Aussage als glaubwürdiger eingestuft, wenn sie mehr Details enthält, plausibler ist, logisch strukturierter ist und der Sender weniger nervös ist. Dies ist daran zu erkennen, dass die Vorzeichen von r_{Per} und r_{Dec} bei den ausgewählten Hinweisreizen gleich sind. r_{Per} bezieht sich dabei auf den Zusammenhang zwischen *wahrgenommener* Täuschung und Lügenindikator, wobei eine positive Korrelation impliziert, dass der Indikator

Experiment 3: Einfluss von unbewussten Urteilszielen und Verarbeitungsmodus auf die Akkuratheit von Glaubwürdigkeitsurteilen unter der Berücksichtigung von Schülermerkmalen

stärker mit wahrgenommener Täuschung assoziiert wird. Dagegen bezieht sich r_{Dec} auf den Zusammenhang zwischen *tatsächlicher* Täuschung und dem Lügenindikator, wobei eine positive Korrelation impliziert, dass der Indikator stärker mit tatsächlicher Täuschung in Zusammenhang steht (siehe Hartwig & Bond, 2011, für eine ausführliche Erläuterung). Des Weiteren wurden vier der sechs Indikatoren (Details, Plausibilität, logische Struktur und Nervosität) u.a. auch deshalb zur Einschätzung ausgewählt, weil es sich bei diesen um sog. *objektive* Lügenindikatoren handelt (DePaulo et al., 2003), d.h. um Verhaltensweisen, bei denen es signifikante Unterschiede zwischen Personen gibt, die die Wahrheit sagen, und Personen, die lügen. Die Indikatoren „Antwortlänge“ und „Blickabwendung“ wurden u.a. ausgewählt, weil es sich dabei um Indikatoren handelt, die allgemein als Lügenindikatoren angesehen werden (Bond, 2006). In Tabelle 1 wird ein Überblick über die ausgewählten Lügenindikatoren gegeben.

Um zu testen, ob es in den 24 verwendeten Schüleraussagen in Experiment 3 signifikante Unterschiede zwischen wahren und erfundenen Aussagen hinsichtlich der objektiv eingeschätzten Hinweisreize gibt, wurden sechs einfaktorielle ANOVAs jeweils mit dem Faktor „Status der Aussage“ (wahr vs. erfunden) und den sechs ausgewählten Indikatoren als abhängigen Variablen durchgeführt. Die Ergebnisse sind ebenfalls in Tabelle 1 dargestellt und deuten darauf hin, dass es signifikante Unterschiede zwischen wahren und erfundenen Aussagen nur in Bezug auf die Antwortlänge gibt und dass – wie erwartet (vgl. DePaulo et al., 2003) – wahre Aussagen länger sind als erfundene Aussagen. Einen marginal signifikanten Unterschied zwischen wahren und

Experiment 3: Einfluss von unbewussten Urteilszielen und Verarbeitungsmodus auf die Akkuratheit von Glaubwürdigkeitsurteilen unter der Berücksichtigung von Schülermerkmalen

erfundenen Aussagen gibt es hinsichtlich des Faktors „Logische Struktur der Aussagen“: So sind wahre Aussagen logisch strukturierter als erfundene Aussagen.

Da die Schülervariable „Antwortlänge“ zum einen eine Variable darstellt, die von Urteilern bei der Glaubwürdigkeitsbeurteilung genutzt wird, und zum anderen eine Variable ist, bei der es signifikante Unterschiede zwischen wahren und erfundenen Aussagen im verwendeten Videomaterial gibt, wird diese in den weiteren Analysen als Variable der Ebene 1 berücksichtigt.

7.2.3 *Versuchsdesign*

In Experiment 3 wurden zur Vorhersage des Kriteriums drei Prädiktoren verwendet: Als Ebene-1-Variable wurde die Schülervariable „Antwortlänge“ in die Analysen aufgenommen, zu den Ebene-2-Variablen zählen „Urteilsvermeidung“ (Vermeidung Urteil wahr vs. Vermeidung Urteil erfunden) und „Hinweisreiz“ (Hinweisreiz wahr vs. Hinweisreiz erfunden).

7.2.4 *Stichprobe*

An Experiment 3 nahmen insgesamt 79 Lehrkräfte teil. Neunzehn Probanden (24%³¹) mussten aus den Analysen aufgrund nicht-gegebener Urteilsbegründungen und fehlender zusätzlicher Begründungen ausgeschlossen werden. Wie bereits erwähnt, wurden die Probanden gebeten, jedes Glaubwürdigkeitsurteil direkt nach jedem Interview kurz zu begründen (*Urteilsbegründung*) und in Abhängigkeit ihres jeweiligen Status bezüglich des

³¹ In Experiment 2 wurden ebenfalls ein Viertel der Probanden (23.9%) aus den weiteren Analysen ausgeschlossen.

Experiment 3: Einfluss von unbewussten Urteilszielen und Verarbeitungsmodus auf die Akkuratheit von Glaubwürdigkeitsurteilen unter der Berücksichtigung von Schülermerkmalen

Faktors „Urteilsvermeidung“ und ihres Urteils eine ausführliche *zusätzliche Begründung* abzugeben. Sofern sie keine Urteilsbegründung ($N = 4$) und/oder zusätzliche Begründung abgaben ($N = 15$), folgten sie nicht den Instruktionen und respektierten somit nicht die Voraussetzungen für die Teilnahme an der Online-Studie. Lediglich eine Person bestand den Instruktionscheck für die Urteilsvermeidung auch beim zweiten Versuch nicht und wurde daher ebenfalls aus den Analysen ausgeschlossen. Die Hälfte der Teilnehmer war weiblich (50.0%). Die Teilnehmer arbeiteten entweder an einer weiterführenden Schule (73.3%) oder einer beruflichen Schule (26.7%). Im Durchschnitt waren die Teilnehmer 40.13 Jahre ($SD = 9.72$) alt und bereits seit 10.40 Jahren ($SD = 8.80$) im Schuldienst.

7.2.5 *Abhängige Variablen und Vorgehen bei der Ergebnisauswertung*

Als abhängige Variablen sind für die folgenden mehrebenenanalytischen Auswertungen zum einen die Wahrscheinlichkeit, eine Aussage als „wahr“ einzustufen (Wahrscheinlichkeit Urteil wahr), und zum anderen die Wahrscheinlichkeit, eine Aussage akkurat einzuschätzen (Wahrscheinlichkeit Urteil richtig) von Bedeutung. Die Hypothesentestung erfolgte mit Hilfe der Software R (<http://www.r-project.org/>). Um die Daten mehrebenenanalytisch mittels R auswerten zu können, wurde der Datensatz zunächst umstrukturiert, indem Variablen (Urteile „erfunden/wahr“) zu Fällen transformiert wurden. Des Weiteren wurden der Datensatz mit den Schülervariablen und der Datensatz mit den Lehrervariablen zusammengeführt (siehe Syntax im Anhang III.b). Die Schätzung der Parameter erfolgte mit Hilfe der RML-Methode

Experiment 3: Einfluss von unbewussten Urteilszielen und Verarbeitungsmodus auf die Akkuratheit von Glaubwürdigkeitsurteilen unter der Berücksichtigung von Schülermerkmalen basierend auf der Laplace-Wahrscheinlichkeitsverteilung (für nähere Informationen siehe Davison, 2003, Kapitel 5).

Um die Ergebnisse von Experiment 2 mit Experiment 3 vergleichen zu können, wurden wie in Experiment 2 zusätzlich die Wahrheitsverzerrung, die durchschnittliche Trefferquote und die Trefferquote für wahre und erfundene Aussagen berechnet und die Hypothesen (soweit möglich) varianzanalytisch getestet. Diese Analysen wurden allerdings nicht mit R, sondern mit Hilfe der Software SPSS durchgeführt.

7.3 Ergebnisse

7.3.1 Hypothesentestung anhand logistischer Mehrebenenanalyse

Wahrscheinlichkeit Urteil wahr. Zur Testung der Hypothese, dass Probanden in der Bedingung „Vermeidung Urteil erfunden“ mit größerer Wahrscheinlichkeit die Aussagen eines Schülers als wahr einstufen als Probanden der Bedingung „Vermeidung Urteil wahr“ (Hypothese 1), wurde eine logistische Mehrebenenanalyse mit der zu vorhersagenden Variable „Urteil erfunden/wahr“ und den festen Effekten „Urteilsvermeidung“ (0 = Urteilsvermeidung wahr, 1 = Urteilsvermeidung erfunden) und „Hinweisreiz“ (0 = Hinweisreiz wahr, 1 = Hinweisreiz erfunden) sowie dem zufälligen Effekt³² „Lehrperson“ durchgeführt (siehe Tabelle 2). Die Ergebnisse werden in

³² Der Effekt „Schülervideo“ wurde aufgrund der geringen Stichprobengröße nicht als zufälliger Effekt in das Modell aufgenommen. Die Ergebnisse von Analysen unter Berücksichtigung dieses zufälligen Effekts sind jedoch ähnlich wie die Ergebnisse, wenn der Effekt nicht berücksichtigt wird.

Experiment 3: Einfluss von unbewussten Urteilszielen und Verarbeitungsmodus auf die Akkuratheit von Glaubwürdigkeitsurteilen unter der Berücksichtigung von Schülermerkmalen

Anlehnung an Field, Miles und Field (2012) sowie Gelman und Hill (2009) interpretiert. Im Vergleich zum Nullmodell (d.h. zum Modell ohne Prädiktoren) ist der Modellfit signifikant besser, wenn die Prädiktoren in das Modell aufgenommen werden ($AIC_{\text{Nullmodell}} = 656.8 > AIC_{\text{Modell}} = 639.1$), $\chi^2(3) = 23.66$, $p < .001$. Der Haupteffekt des Intercepts und des Hinweisreizes waren nicht signifikant, $ps > .10$. Wie erwartet ist der Haupteffekt für die Urteilsvermeidung signifikant, $b = 0.83$, $z(417) = 3.01$, $p = .003$. Die Odds Ratio³³, ein Urteil „wahr“ (im Vergleich zu „erfunden“) abzugeben, ist für eine Lehrkraft, die den Hinweisreiz „erfunden“ erhalten hat, 1.97mal größer, wenn diese Lehrkraft das Urteil „erfunden“ (im Vergleich zu „wahr“) vermeidet. Ebenso ist die Odds Ratio, ein Urteil „wahr“ (im Vergleich zu „erfunden“) abzugeben, für eine Lehrkraft, die den Hinweisreiz „wahr“ erhalten hat, 2.29mal größer, wenn diese Lehrkraft das Urteil „erfunden“ (im Vergleich zu „wahr“) vermeidet. Diese Befunde unterstützen Hypothese 1 und deuten darauf hin, dass Personen mit einer Urteilsvermeidung „erfunden“ eine höhere Wahrscheinlichkeit haben, das Urteil „wahr“ abzugeben, als Personen mit der Urteilsvermeidung „wahr“. Keine weiteren Effekte waren signifikant, $ps > .10$.

³³ Die Odds Ratio wurden nach Field (2009, S. 288) berechnet.

Tabelle 2

Geschätzte Regressionsgewichte für feste Effekte (Urteilsvermeidung: 0 = wahr, 1 = erfunden; Hinweisreiz 0 = wahr, 1 = erfunden) bei logistischer Mehrebenenanalyse auf Kriteriumsvariable „Urteil erfunden/wahr“ (0 = erfunden, 1 = wahr)

	<i>b</i>	SE <i>b</i>	z-Wert	<i>p</i>
Intercept	0.154	0.197	0.784	.43
Urteilsvermeidung	0.830	0.275	3.014	.003
Hinweisreiz	-0.419	0.256	-1.637	.101
Urteilsvermeidung x Hinweisreiz	-0.150	0.387	-0.388	.698

Wahrscheinlichkeit richtiges Urteil. Um zu testen, ob Probanden in der Bedingung mit urteilszielinkonsistentem Hinweisreiz, d.h. in der Bedingung „Vermeidung Urteil erfunden“ und „Hinweisreiz erfunden“ bzw. in der Bedingung „Vermeidung Urteil wahr“ und „Hinweisreiz wahr“, mit größerer Wahrscheinlichkeit die Aussagen eines Schülers richtig als wahr im Vergleich zu erfunden einstufen als Probanden mit urteilszielkonsistentem Hinweisreiz (Hypothese 2), wurde ebenfalls eine logistische Mehrebenenanalyse mit der zu vorhersagenden Variable „Urteil falsch/richtig“ und den festen Effekten „Urteilsvermeidung“ und „Hinweisreiz“ sowie dem zufälligen Effekt „Lehrperson“ durchgeführt (siehe Tabelle 3). Hierbei ist anzumerken, dass der Intercept Werte zwischen 0 (entspricht „Urteil falsch“) und 1 (entspricht „Urteil richtig“) annehmen kann. Im Vergleich zum Nullmodell ist der Modellfit signifikant besser, wenn die Prädiktoren in das Modell aufgenommen werden ($AIC_{\text{Nullmodell2}} = 654.6 > AIC_{\text{Modell2}} = 652.6$), $\chi^2(3) = 8.06$, $p = .044$. Wie erwartet ist der Interaktionseffekt zwischen „Urteilsvermeidung“ und „Hinweisreiz“

Experiment 3: Einfluss von unbewussten Urteilszielen und Verarbeitungsmodus auf die Akkuratheit von Glaubwürdigkeitsurteilen unter der Berücksichtigung von Schülermerkmalen

signifikant, $b = 0.782$, $z(417) = 2.03$, $p = .043$. Die Odds Ratio³⁴, ein richtiges Urteil (im Vergleich zu einem falschen Urteil) abzugeben, ist für eine Lehrkraft der Bedingung „Vermeidung Urteil wahr“ 1.04mal größer, wenn diese Lehrkraft den urteilszielinkonsistenten Hinweisreiz „wahr“ (im Vergleich zu „erfunden“) erhalten hat. Ebenso ist die Odds Ratio, ein richtiges Urteil (im Vergleich zu einem falschen Urteil) abzugeben, für eine Lehrkraft der Bedingung „Vermeidung Urteil erfunden“ 2.09mal größer, wenn diese Lehrkraft den urteilszielinkonsistenten Hinweisreiz „erfunden“ (im Vergleich zu „wahr“) erhalten hat. Der Haupteffekt der Urteilsvermeidung war marginal signifikant, $p = .059$, der Haupteffekt des Hinweisreizes war nicht signifikant, $p = .87$.

Um den Effekt des Interaktionsterms aufzuschlüsseln, wurden getrennte Analysen für die beiden Arten der Urteilsvermeidung durchgeführt. Eine Analyse der Lehrpersonen der Bedingung „Vermeidung Urteil erfunden“ zeigte einen signifikanten Haupteffekt des Hinweisreizes, $b = 0.739$, $z(195) = 2.60$, $p = .009$. Dies deutet darauf hin, dass Lehrpersonen der Bedingung „Vermeidung Urteil erfunden“ eine größere Wahrscheinlichkeit haben, ein richtiges Urteil zu fällen, wenn sie einen urteilszielinkonsistenten „Hinweisreiz erfunden“ im Vergleich zum urteilszielkonsistenten Hinweisreiz „wahr“ erhalten hatten. Bei Lehrpersonen der Bedingung „Vermeidung Urteil wahr“ gab es keine signifikanten Unterschiede zwischen den beiden Hinweisreizen, $p = .87$.

³⁴ Die Odds Ratio wurden nach Field (2009, S. 288) berechnet.

Tabelle 3

Geschätzte Regressionsgewichte für feste Effekte (Urteilsvermeidung: 0 = wahr, 1 = erfunden; Hinweisreiz 0 = wahr, 1 = erfunden) bei logistischer Mehrebenenanalyse auf Kriteriumsvariable „Urteil falsch/richtig“ (0 = falsch, 1 = richtig)

	<i>b</i>	SE <i>b</i>	z-Wert	<i>p</i>
Intercept	0.470	0.202	2.332	.020
Urteilsvermeidung	-0.499	0.265	-1.887	.059
Hinweisreiz	-0.043	0.261	-0.163	.870
Urteilsvermeidung x Hinweisreiz	0.782	0.386	2.026	.043

Cross-Level-Interaktion. Laut Hypothese 3 wird eine Dreifachinteraktion zwischen den Faktoren „Urteilsvermeidung“, „Hinweisreiz“ und „Antwortlänge“ auf das Urteil „erfunden/wahr“ erwartet. Insbesondere wird angenommen, dass Lehrpersonen in urteilsziel~~in~~konsistenten Bedingungen durch die Länge der Antwort des Schülers in ihrem Urteil beeinflusst werden. Um die Interpretation der Variable „Antwortlänge“ zu erleichtern, wurde die Antwortlänge um den Gesamtmittelwert zentriert (sog. Grand Mean Centering, Twisk, 2011). Der Wert 0 entspricht damit der mittleren Antwortlänge, der Wert 1 steht für Schüleraussagen, deren Antwortlänge eine Einheit über dem Gesamtmittelwert liegt. Die Ergebnisse sind in Tabelle 4 dargestellt.

Tabelle 4

Geschätzte Regressionsgewichte für feste Effekte (Urteilsvermeidung: 0 = wahr, 1 = erfunden; Hinweisreiz 0 = wahr, 1 = erfunden) bei logistischer Mehrebenenanalyse auf Kriteriumsvariable „Urteil erfunden/wahr“ (0 = erfunden, 1 = wahr)

	<i>b</i>	SE <i>b</i>	z-Wert	<i>p</i>
Intercept	0.148	0.198	0.746	.45
Urteilsvermeidung	0.844	0.277	3.044	.002
Hinweisreiz	-0.430	0.258	-1.664	.096
Antwortlänge	0.003	0.002	1.205	.228
Urteilsvermeidung x Hinweisreiz	0.094	0.414	0.225	.82
Urteilsvermeidung x Antwortlänge	-0.004	0.003	-1.192	.23
Hinweisreiz x Antwortlänge	0.001	0.003	0.006	.99
Urteilsvermeidung x Hinweisreiz x Antwortlänge	0.009	0.005	1.938	.053

Im Vergleich zum Nullmodell ist der Modellfit signifikant besser, wenn die Prädiktoren in das Modell aufgenommen werden ($AIC_{\text{Nullmodell3}} = 656.8 > AIC_{\text{Modell3}} = 635.0$), $\chi^2(7) = 35.84$, $p < .001$. Entsprechend Hypothese 3 zeigen die Befunde eine marginal signifikante Dreifachinteraktion zwischen den Faktoren „Urteilsvermeidung“, „Hinweisreiz“ und „Antwortlänge“, $b = 0.009$, $z(413) = 0.78$, $p = .053$.

Um den Effekt der Dreifachinteraktion aufzuschlüsseln, wurden getrennte Analysen für alle vier Bedingungen (Urteilsvermeidung x Hinweisreiz) durchgeführt. Die Ergebnisse zeigen, dass nur für Lehrpersonen der Bedingung „Vermeidung Urteil erfunden“ und einem urteilszielinkonsistentem „Hinweisreiz erfunden“ die Wahrscheinlichkeit für ein

„wahr“-Urteil steigt, wenn die Antwortlänge eine Einheit über dem
Durchschnitt liegt, $b = 0.008$, $z(76) = 2.59$, $p = .010$. In den drei übrigen
Bedingungen hatte die Antwortlänge keinen signifikanten Effekt auf die
Wahrscheinlichkeit Urteil „wahr“, alle $ps > .13$.

7.3.2 *Hypothesentestung anhand von ANOVAs*

Die folgenden Analysen wurden mit SPSS durchgeführt, um eine
Vergleichbarkeit der Ergebnisse von Experiment 3 mit Experiment 2 erzielen zu
können.

Urteilsverzerrung. Im Durchschnitt wurden $M = 4.57$ ($SD = 1.63$)
Aussagen als wahr eingestuft. Dieser Mittelwert ist signifikant größer als 4,
 $t(60) = 2.69$, $p = .009$. Um zu testen, ob die Manipulation des Urteilsziels dazu
führte, dass die Teilnehmer die Aussagen als tendenziell wahr oder erfunden
einstufen und somit die Urteilsverzerrung beeinflusst wurde, wurde eine
ANOVA mit der Wahrheitsverzerrung als abhängiger Variable und der
Urteilsvermeidung und dem Hinweisreiz als unabhängigen Variablen
durchgeführt. Die Ergebnisse zeigten den erwarteten signifikanten Haupteffekt
des Faktors „Urteilsvermeidung“ auf die Wahrheitsverzerrung, $F(1, 56) = 15.53$,
 $p < .001$, $\eta^2 = .217$. Somit stuften die Probanden der Bedingung „Vermeidung
Urteil erfunden“ signifikant häufiger die Aussagen als wahr ein ($M = 5.43$, $SD =$
 1.43) und die Probanden der Bedingung „Vermeidung Urteil wahr“ diese
häufiger als erfunden ein ($M = 3.81$, $SD = 1.42$). Des Weiteren zeigte sich ein
signifikanter Haupteffekt des Faktors „Hinweisreiz“: Lehrpersonen, die einen
Hinweisreiz „wahr“ erhalten hatten, beurteilten die Aussagen eher als wahr (M
 $= 5.17$, $SD = 1.53$) als Lehrpersonen, die einen Hinweisreiz „erfunden“ erhalten

Experiment 3: Einfluss von unbewussten Urteilszielen und Verarbeitungsmodus auf die Akkuratheit von Glaubwürdigkeitsurteilen unter der Berücksichtigung von Schülermerkmalen hatten ($M = 3.97$, $SD = 1.52$). Der Effekt der Interaktion war nicht signifikant ($F < 1$).

Trefferquote. Die durchschnittliche Trefferquote der Probanden lag bei 58.75% ($SD = 16.16$) und damit leicht über dem Zufallsniveau³⁵ 50%, $t(60) = 4.19$, $p < .001$. Somit waren die Probanden leicht besser als der Zufall im Erkennen von wahren und erfundenen Aussagen. Dieses Ergebnis deckt sich mit bisherigen Befunden (vgl. C. F. Bond & DePaulo, 2006). Des Weiteren war die Trefferquote nicht signifikant mit Geschlecht, Alter oder beruflicher Erfahrung (in Jahren) korreliert, alle $ps > .54$. Um den Effekt des Faktors „Urteilsvermeidung“ und urteilsziel(in)konsistentem Hinweisreiz zu testen, wurde eine ANOVA mit „Trefferquote“ als abhängiger Variable sowie „Urteilsvermeidung“ und „Hinweisreiz“ als unabhängigen Variablen durchgeführt. Die Ergebnisse zeigen, dass der erwartete Interaktionseffekt von Urteilsvermeidung und Hinweisreiz auf die Trefferquote signifikant war, $F(1, 56) = 5.47$, $p = .023$, $\eta^2 = .089$. Somit waren die Probanden in der Bedingung „Vermeidung Urteil erfunden“ signifikant besser im Unterscheiden von Wahrheit und Lüge nachdem sie einen urteilszielinkonsistenten Lügenhinweisreiz erhalten hatten ($M = 67.05\%$, $SD = 21.12$) als nach dem Erhalt des urteilszielkonsistenten Wahrheitshinweisreizes ($M = 49.26\%$, $SD = 14.97$), $F(1, 56) = 9.12$, $p = .004$ (Simple-Effect-Analysen nach Field, 2009). Probanden in der Bedingung „Vermeidung Urteil wahr“ waren nicht signifikant akkurater,

³⁵ Wie bereits oben erläutert wird unter Zufallsniveau die Ratewahrscheinlichkeit verstanden, wenn 50% wahre und 50% erfundene Aussagen zur Beurteilung vorgelegt werden (vgl. Vrij, 2008, S. 141).

Experiment 3: Einfluss von unbewussten Urteilszielen und Verarbeitungsmodus auf die Akkuratheit von Glaubwürdigkeitsurteilen unter der Berücksichtigung von Schülermerkmalen

nachdem sie den urteilszielinkonsistenten (Wahrheits-)Hinweisreiz erhalten hatten ($M = 61.54$, $SD = 11.93$) als nach Erhalt des urteilszielkonsistenten (Lügen-)Hinweisreizes ($M = 60.53$, $SD = 13.35$), $F < 1$. Weitere Simple-Effect-Analysen deuten darauf hin, dass nach Erhalt des „Hinweisreiz wahr“ Probanden in der urteilszielkonsistenten Bedingung „Vermeidung Urteil erfunden“ signifikant schlechter im Erkennen von Wahrheit und Lüge waren als Probanden in der urteilszielinkonsistenten Bedingung „Vermeidung Urteil wahr“. Nach Erhalt eines „Hinweisreiz erfunden“ gab es keine signifikanten Unterschiede zwischen den beiden Urteilsvermeidungsbedingungen, $p = .26$. Weiterhin war der Haupteffekt des Hinweisreizes signifikant, $F(1, 56) = 4.35$, $p = .042$, $\eta^2 = .072$, und Probanden akkurater, nachdem sie den „Hinweisreiz erfunden“ ($M = 62.92$, $SD = 16.57$) im Vergleich zum „Hinweisreiz wahr“ erhalten hatten ($M = 54.58$, $SD = 14.86$). Der Interaktionseffekt ist in Abbildung 8 dargestellt. Der Haupteffekt der Urteilsvermeidung war nicht signifikant ($p = .48$).

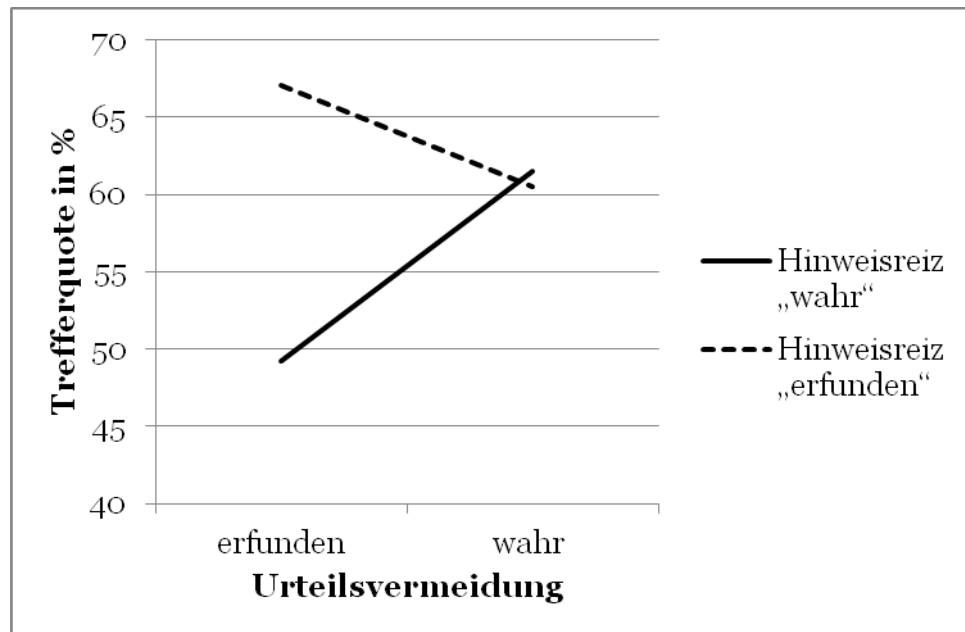


Abbildung 8. Trefferquote in Prozent in Abhängigkeit der Urteilsvermeidung und des Hinweisreizes in Experiment 3.

Trefferquote bei wahren und erfundenen Aussagen. Um zu überprüfen, ob Probanden besser im Erkennen von wahren oder erfundenen Aussagen waren, wurde eine ANOVA mit Messwiederholung durchgeführt mit „Valenz der Aussage“ (Trefferquote bei wahren und erfundenen Aussagen) als Innersubjektfaktor sowie „Urteilsvermeidung“ und „Hinweisreiz“ als Zwischensubjektfaktoren. Wie erwartet waren die Probanden signifikant besser im Erkennen von wahren Aussagen ($M = 65.83\%$, $SD = 25.20$) als im Erkennen von erfundenen Aussagen ($M = 51.67\%$, $SD = 26.79$), $F(1, 56) = 11.15$, $p = .002$, $\eta^2 = .17$. Die durchschnittliche Trefferquote für wahre Aussagen unterscheidet sich signifikant vom Zufallsniveau 50%, $t_{\text{wahr}}(59) = 4.87$, $p_{\text{wahr}} < .001$, die Trefferquote für erfundene Aussagen liegt jedoch auf Zufallsniveau, $p = .63$. Dieser Haupteffekt des Innersubjektfaktors wurde – wie erwartet – qualifiziert durch eine signifikante Interaktion zwischen „Valenz der Aussage“ und „Urteilsvermeidung“, $F(1, 56) = 15.53$, $p < .001$, $\eta^2 = .22$. Simple-Effect-

Experiment 3: Einfluss von unbewussten Urteilszielen und Verarbeitungsmodus auf die Akkuratheit von Glaubwürdigkeitsurteilen unter der Berücksichtigung von Schülermerkmalen

Analysen (vgl. Field, 2009, siehe Syntax III.c im Anhang) zeigen, dass Probanden der Bedingung „Vermeidung Urteil erfunden“ signifikant besser im Erkennen von wahren Aussagen waren ($M = 74.11\%$, $SD = 25.89$) als Probanden der Bedingung „Vermeidung Urteil wahr“ ($M = 58.59\%$, $SD = 22.55$), $F(1, 58) = 6.16$, $p = .016$, $\eta^2 = .096$. Dagegen waren Probanden der Bedingung „Vermeidung Urteil wahr“ signifikant besser im Erkennen von erfundenen Aussagen ($M = 63.28\%$, $SD = 21.05$) als Probanden der Bedingung „Vermeidung Urteil erfunden“ ($M = 38.39\%$, $SD = 26.77$), $F(1, 58) = 16.22$, $p < .001$, $\eta^2 = .22$. Die Zweifachinteraktion zwischen den Faktoren „Valenz der Aussage“ und „Hinweisreiz“ war ebenfalls signifikant, $F(1, 56) = 6.42$, $p = .014$, $\eta^2 = .103$. Die Dreifachinteraktion war nicht signifikant, $F < 1$.

7.4 Diskussion Experiment 3

Wie auch in Experiment 2 wurde mittels Experiment 3 getestet, inwiefern Lehrpersonen die Tendenz haben, Aussagen von Schülern als „wahr“ einzustufen (Hypothese 1), und ob urteilszielinkonsistente Hinweisreize die Trefferquote erhöhen (Hypothese 2). Diese Hypothesen wurden zum einen anhand einer logistischen Mehrebenenanalyse getestet und es wurde die Wahrscheinlichkeit, ein Urteil „wahr“ abzugeben (Hypothese 1) und ein richtiges Urteil zu fällen (Hypothese 2), geschätzt. Zum anderen wurden die beiden Hypothesen parallel zu Experiment 2 anhand von ANOVAs getestet und der Einfluss von Urteilsvermeidung und Hinweisreiz auf die Urteilsverzerrung und die durchschnittliche Trefferquote untersucht. Sowohl die Ergebnisse der Mehrebenenanalyse als auch die mittels ANOVAs deuten darauf hin, dass Lehrpersonen der Bedingung „Vermeidung Urteil erfunden“ (vs. „wahr“) einer

Wahrheitsverzerrung unterliegen bzw. eine höhere Chance haben, die Schüleraussagen als „wahr“ im Vergleich zu „erfunden“ zu beurteilen. Diese Ergebnisse unterstützen Hypothese 1.

Neben dem erwarteten Haupteffekt des Faktors „Urteilsvermeidung“ zeigte die ANOVA jedoch noch einen Haupteffekt des Faktors „Hinweisreiz“ auf das Glaubwürdigkeitsurteil. Dies wurde aus theoretischer Sicht nicht erwartet; auch aufgrund des Vortests des Hinweisreizes (siehe Experiment 2) wurde kein Haupteffekt erwartet. Der Vortest wurde durchgeführt, um zu untersuchen, ob Lehrpersonen sich nach Erhalt des Hinweisreizes der Gruppennorm anpassen und mit der Urteilstendenz der Gruppe konform urteilen. Es wurde erwartet, dass der Hinweisreiz keine Konformität mit der Gruppennorm zur Folge hat, da der Hinweisreiz die Information enthielt, dass die Probanden der kurzen Voruntersuchung keinen zusätzlichen bzw. ausführlichen Bericht nach Treffen ihren Glaubwürdigkeitsurteilen schreiben mussten. Diese Information gab den Probanden die Möglichkeit, die Antworttendenz der Probanden der kurzen Voruntersuchung vernachlässigen zu können, da die Bedingungen, unter denen die Probanden der kurzen Voruntersuchung standen (d.h. sie mussten keine ausführliche Begründung schreiben), nicht gleichwertig mit den eigenen Bedingungen waren. Die Ergebnisse des Vortests zeigten, dass kein Konformitätsdruck durch die bloße Darbietung des Hinweisreizes erzeugt wurde. Das Auftreten eines Haupteffekts des Faktors „Hinweisreiz“ in Experiment 3 (und nicht in Experiment 2) könnte dadurch erklärt werden, dass sich die verwendeten Videoaufnahmen in Experiment 2 und 3 unterschieden. So lag die *durchschnittliche* Länge der Aufnahmen sowohl in Experiment 2 als

Experiment 3: Einfluss von unbewussten Urteilszielen und Verarbeitungsmodus auf die Akkuratheit von Glaubwürdigkeitsurteilen unter der Berücksichtigung von Schülermerkmalen

auch in Experiment 3 bei ca. 2 Minuten, die Standardabweichung der Videoaufnahmen war jedoch bei den Videoaufnahmen aus Experiment 3 ungefähr viermal so groß ($SD = 0.64$) wie bei denen aus Experiment 2 ($SD = 0.15$). Die Unterschiedlichkeit in der Dauer der Aufnahmen wird auch durch die minimale und maximale Dauer deutlich (Experiment 2: Min = 1.96, Max = 2.43; Experiment 3: Min = 1.2, Max = 4.08). Dadurch, dass sich die Videoaufnahmen in Experiment 3 hinsichtlich ihrer Dauer stark von denen in Experiment 2 unterschieden, hatten die Probanden weniger Vergleichsmöglichkeiten zwischen den Videoaufnahmen: So ist ein Vergleich z.B. hinsichtlich der Anzahl der Details leichter, wenn Videoaufnahmen gleicher Dauer sind. Dies könnte dazu geführt haben, dass die Aussagen als weniger eindeutig wahrgenommen wurden. In nicht eindeutigen Situationen tendieren Personen dazu, informationalem sozialem Einfluss zu folgen und sich der Gruppennorm anzupassen. Somit kann davon ausgegangen werden, dass in Experiment 3 informationaler sozialer Einfluss zum Tragen und es dadurch zur Anpassung an die Gruppennorm kam (vgl. Aronson et al., 2008, zu Bedingungen von Konformität, Kapitel 8.3).

Dass dieser Haupteffekt des Faktors „Hinweisreiz“ auf das Glaubwürdigkeitsurteil lediglich anhand der ANOVA, nicht jedoch anhand der Mehrebenenanalyse aufgedeckt werden konnte, kann dadurch erklärt werden, dass bei der Mehrebenenregression der Prädiktor nicht aggregiert (Summe Urteil „wahr“) war und dadurch die Werte akkurater geschätzt werden konnten.

Hypothese 2 kann ebenfalls sowohl anhand der ANOVA als auch anhand der Mehrebenenanalyse unterstützt werden; die Interaktion zwischen den

Faktoren „Urteilsvermeidung“ und „Hinweisreiz“ war bei beiden Analyseverfahren signifikant. Dies deutet darauf hin, dass Lehrpersonen mit einem urteilszielinkonsistenten Hinweisreiz besser Lüge und Wahrheit unterscheiden konnten. Weitere Analysen (Simple-Effect-Analyse und getrennte Mehrebenenregressionsanalysen) deuten jedoch darauf hin, dass der Unterschied zwischen urteilszielkonsistentem und -inkonsistentem Hinweisreiz lediglich für die Bedingung „Vermeidung Urteil erfunden“ signifikant ist. Dies könnte dadurch erklärt werden, dass in der Bedingung „Vermeidung Urteil wahr“ der inkonsistente „Hinweisreiz wahr“ keine „Bedrohung“ darstellte, sondern lediglich die „natürliche“ Urteilsverzerrung (vgl. Hurst-Wagner & Oswald, 2011) der Probanden unterstützte. So konnten bisherige Studien zeigen, dass Lehrkräfte eher einer Wahrheitsverzerrung unterliegen (Vrij et al., 2006) und dass diese Verzerrung sogar mit der Erfahrung als Lehrkraft zunimmt (Reinhard, Dickhäuser, et al., 2011). Die Annahme, dass der Hinweisreiz keine „Bedrohung“ darstellte, wird durch den Haupteffekt des Faktors „Hinweisreiz“ auf das Glaubwürdigkeitsurteil unterstützt (siehe Ergebnisse ANOVA): Probanden stufen die Aussagen von Schülern – unabhängig von ihrer Urteilsvermeidung – eher als wahr ein, nachdem sie einen „Hinweisreiz wahr“ erhalten haben, und eher als erfunden, wenn sie einen „Hinweisreiz erfunden“ erhalten haben.

Die Ergebnisse der ANOVA zeigen einen nicht erwarteten Haupteffekt des Hinweisreizes auf die Trefferquote: Probanden hatten eine höhere Trefferquote, nachdem sie den Hinweisreiz „erfunden“ im Vergleich zu „wahr“ erhalten hatten – unabhängig vom Faktor „Urteilsvermeidung“. Dies kann zum

Experiment 3: Einfluss von unbewussten Urteilszielen und Verarbeitungsmodus auf die Akkuratheit von Glaubwürdigkeitsurteilen unter der Berücksichtigung von Schülermerkmalen

einen durch den Haupteffekt des Faktors „Hinweisreiz“ auf das Glaubwürdigkeitsurteil und zum anderen durch die „natürliche“ Tendenz der Lehrpersonen, Aussagen als „wahr“ einzustufen, und den Wahrheitseffekt (d.h. tatsächlich wahre Aussagen werden besser als solche erkannt) erklärt werden. Über alle Bedingungen hinweg unterlagen Lehrpersonen einer Wahrheitsverzerrung und haben somit eine höhere Wahrscheinlichkeit, wahre Aussagen akkurat als solche zu erkennen. Da der Faktor „Hinweisreiz“ sich aber signifikant auf das Glaubwürdigkeitsurteil auswirkt und Lehrpersonen nach Erhalt eines „Hinweisreiz erfunden“ eher die Aussage als „erfunden“ einstufen, reduziert sich die Wahrscheinlichkeit, wahre Aussagen akkurat einzuschätzen, und die Wahrscheinlichkeit, erfundene Aussagen akkurat einzuschätzen, erhöht sich – gleichzeitig erhöht sich auch die allgemeine Trefferquote.

Dass dieser Haupteffekt des Faktors „Hinweisreiz“ auf die Trefferquote anhand der ANOVA, nicht jedoch anhand der Mehrebenenanalyse aufgedeckt werden konnte, kann ebenfalls dadurch erklärt werden, dass bei der Mehrebenenregression der Prädiktor nicht aggregiert (Summe Urteil „wahr“) war und dadurch die Werte akkurater geschätzt werden konnten.

Die Ergebnisse der Mehrebenenregression mit Berücksichtigung der Schülervariable „Antwortlänge“ unterstützen Hypothese 3 und deuten darauf hin, dass die Annahme, dass systematische Verarbeitung zu einer Nutzung verbaler Informationen führt, zutreffend ist. So zeigt sich, dass Probanden der Bedingung „Vermeidung Urteil erfunden“ eine höhere Wahrscheinlichkeit haben, die Aussagen als „wahr“ einzustufen, wenn sie einen urteilszielinkonsistenten „Hinweisreiz erfunden“ im Vergleich zu „wahr“

Experiment 3: Einfluss von unbewussten Urteilszielen und Verarbeitungsmodus auf die Akkuratheit von Glaubwürdigkeitsurteilen unter der Berücksichtigung von Schülermerkmalen erhalten hatten. Dies kann dadurch erklärt werden, dass mit einer höheren Antwortlänge die Annahme einhergeht, dass Personen die Wahrheit sagen (vgl. Bond, 2006). Dieser Hinweisreiz (Antwortlänge) scheint wie erwartet nur von Lehrpersonen der Bedingung „Vermeidung Urteil erfunden“ genutzt zu werden und einen urteilszielinkonsistenten im Vergleich zu einem urteilszielkonsistenten Hinweisreiz erhalten haben. Allerdings trifft dies nicht auf Probanden der Bedingung „Vermeidung Urteil wahr“ und mit einem inkonsistenten Hinweisreiz zu. Dies könnte dadurch erklärt werden, dass der Hinweisreiz „Antwortlänge“ als „Wahrheitsindikator“ (vgl. Bond, 2006) angesehen wird und somit genutzt wird, um eine Aussage als „wahr“ einzustufen – nicht jedoch um eine Aussage als „erfunden“ zu beurteilen. Es kann angenommen werden, dass „Lügenindikator“ (z.B. die Nervosität einer Person oder die Inkonsistenz einer Aussage) genutzt werden, um eine Aussage als „erfunden“ einzustufen. Somit kann Hypothese 3 als teilweise bestätigt angesehen werden.

Neben der marginal signifikanten Dreifachinteraktion zeigte die Mehrebenenregression mit der Schülervariable „Antwortlänge“ einen signifikanten Haupteffekt des Faktors „Urteilsvermeidung“. Dies unterstützt Hypothese 1: Lehrpersonen der Bedingung „Vermeidung Urteil erfunden“ haben eine höhere Wahrscheinlichkeit, Aussagen als „wahr“ einzustufen, als Lehrpersonen der Bedingung „Vermeidung Urteil wahr“.

Insgesamt zeigt sich, dass durch eine Mehrebenenregression mit nicht-aggregierter vorherzusagenden Variable teilweise unterschiedliche Ergebnisse erzielt werden als mit einer ANOVA mit aggregierter abhängiger Variable. So

Experiment 3: Einfluss von unbewussten Urteilszielen und Verarbeitungsmodus auf die Akkuratheit von Glaubwürdigkeitsurteilen unter der Berücksichtigung von Schülermerkmalen

legen die Ergebnisse der ANOVA einen Haupteffekt des Faktors „Hinweisreiz“ sowohl auf das Glaubwürdigkeitsurteil als auch auf die Trefferquote nahe. Diese Befunde konnten nicht durch die Mehrebenenregression unterstützt werden. Eine mögliche Erklärung hierfür könnte sein, dass durch die Aggregation der Daten Effekte erzeugt werden, die tatsächlich nicht vorhanden sind. Daher kann angenommen werden, dass die Auswertung von Messwiederholungsdaten anhand einer Mehrebenenregression akkurater ist, als eine Aggregation von Daten mit anschließender varianzanalytischer Auswertung.

Ein weiterer Vorteil von Mehrebenenregressionen und der Hauptgrund für die Nutzung dieser Analysemethode in Experiment 3 ist die Möglichkeit, Variablen erster Ebene, d.h. Schülervariablen, bei den Analysen berücksichtigen zu können. Da in der vorliegenden Arbeit die Stichprobe relativ gesehen gering war, konnte lediglich eine einzige Schülervariable mit in die Analysen aufgenommen werden. In zukünftigen Arbeiten könnte eine größere Stichprobe es ermöglichen, dass mehrere Schülervariablen gleichzeitig berücksichtigt werden und die Effekte von „Wahrheitsindikatoren“ (z.B. Antwortlänge) und „Lügenindikatoren“ (z.B. Nervosität) gleichzeitig getestet werden.

8 Abschlussdiskussion

8.1 Effekt von unbewussten Urteilszielen und Verarbeitungsmodus auf Glaubwürdigkeitsbeurteilung von Schülern

8.1.1 *Einfluss auf die Glaubwürdigkeitsbeurteilung*

Betrachtet man die Ergebnisse der durchgeführten Experimente, so zeigt sich, dass unbewusste Urteilsziele einen direkten Einfluss auf das Glaubwürdigkeitsurteil bzw. die Wahrscheinlichkeit, Aussagen als „wahr“ einzustufen, haben und dass Schülervariablen das Glaubwürdigkeitsurteil beeinflussen.

Die Ergebnisse von Experiment 1 und Pilotstudie 2 zeigen, dass ein Urteilsziel „wahr“ die Wahrscheinlichkeit erhöht, dass Lehrkräfte den Aussagen eines Schülers glauben und diese als „wahr“ einstufen. Wichtig ist hierbei, dass die Urteile auf einer verzerrten Verarbeitung der Schüleraussagen beruhten: So zeigte sich in beiden Experimenten, dass mehr Argumente generiert wurden, welche die Unschuld des Schülers unterstützten, wenn Lehrkräfte ein Urteilsziel „wahr“ verfolgten, und mehr Argumente genannt wurden, die für die Schuld des Schülers sprachen, wenn die Lehrkräfte ein Urteilsziel „erfunden“ verfolgten. Die Urteilsziele sind dabei – wie Pilotstudie 1 und 2 zeigen – unbewusst verankert.

Wie oben dargestellt, gibt es Hinweise dafür, dass es sich bei Urteilzielen tatsächlich um „Ziele“ handelt. Getestet wurde dies im Rahmen dieser Arbeit jedoch nicht. Wie bereits in Kapitel 1.6.1 erläutert sind Ziele etwas, das ein Individuum zu erreichen versucht (Locke et al., 1981). Ausgehend von dieser

Definition kann angenommen werden, dass Ziele dem Handelnden bewusst und zugänglich sind und somit durch Erfragen erfasst werden können (z.B. durch die Zustimmung zur Aussage „Nachfolgend versuche ich möglichst häufig das Urteil ‚wahr‘ abzugeben.“). So kann davon ausgegangen werden, dass Lehrkräfte mit einem Urteilsziel „wahr“ diese Aussage eher zustimmen als Lehrkräfte mit einem Urteilsziel „erfunden“. „Ziele“ können jedoch auch unbewusst und dem Handelnden unzugänglich sein (vgl. Uleman, 1996, 1999). Dann wäre eine direkte Befragung wenig sinnvoll und hilfreich. Hier könnte eine indirekte Erfassung z.B. durch den Implicit Association Test (Greenwald, McGhee, & Schwartz, 1998) zielführender sein (vgl. auch Kapitel 8.3, Ausblick).

Die Ergebnisse von Experiment 1 deuten darüber hinaus darauf hin, dass angehende Lehrkräfte Schüleraussagen auf eine verzerrte *systematische* Art und Weise verarbeiten, wenn sie einen urteilszielinkonsistenten Hinweisreiz erhalten haben und im Glauben waren, dass Personen der gleichen sozialen Gruppe die Aussagen des Schülers entgegen dem induzierten Urteilsziel beurteilt hatten. Dies wird dadurch erklärt, dass ein Hinweisreiz, der den eigenen Zielen widerstrebt, als Bedrohung wahrgenommen wird; diese Bedrohung wird versucht abzuwenden, indem es zu einer systematischen, kognitiv aufwendigeren Verarbeitung der urteilsrelevanten Informationen kommt (vgl. Giner-Sorolila & Chaiken, 1997).

Experiment 2 und 3 stützen diese Befunde und zeigen, dass Lehrpersonen entsprechend ihres Urteilsziels einer Urteilsverzerrung unterliegen. Sie deuten darauf hin, dass Lehrpersonen einer Wahrheitsverzerrung unterliegen, wenn sie informiert wurden, dass sie für

jedes Urteil „erfunden“ eine ausführliche Begründung abgeben müssen und daher das Urteilsziel „wahr“ verfolgen. Dagegen zeigen Lehrpersonen eine Lügenverzerrung, wenn sie das Urteilsziel „erfunden“ verfolgen und jedes Mal ein Urteil ausführlich begründen müssen, wenn sie die Aussagen als „wahr“ eingestuft haben. Diese Befunde konnten sowohl anhand einer ANOVA als auch einer Mehrebenenregression aufgedeckt werden.

Eine weitere Variable, die das Glaubwürdigkeitsurteil beeinflusst, ist die Antwortlänge der Aussagen der Schüler in Abhängigkeit vom Urteilsziel und des Verarbeitungsmodus. Die Ergebnisse der Mehrebenenregression in Experiment 3 zeigen, dass Lehrpersonen überdurchschnittlich lange Schüleraussagen mit größerer Wahrscheinlichkeit als „wahr“ einstufen, wenn sie das Urteilsziel „wahr“ verfolgten und zusätzlich einen urteilszielinkonsistenten (im Vergleich zu einem urteilszielkonsistenten) Hinweisreiz erhalten haben. Dies kann dadurch erklärt werden, dass Lehrpersonen mit einem urteilszielinkonsistenten Hinweisreiz systematisch verarbeiteten und somit verbale Hinweise wie die Antwortlänge zur Bildung des Glaubwürdigkeitsurteils nutzen. Inwiefern nonverbale Informationen oder Hinweise, die subjektiv als Indikatoren für eine Lüge angesehen werden (z.B. die Nervosität eines Senders oder die Inkonsistenz in einer Aussage), Einfluss haben, wurde hier nicht thematisiert.

8.1.2 Einfluss auf die Akkuratheit von Glaubwürdigkeitsbeurteilungen

Die Befunde der durchgeführten Experimente deuten darauf hin, dass unbewusste Urteilsziele und Verarbeitungsmodus einen Einfluss auf die Akkuratheit von Glaubwürdigkeitsurteilen haben und dass

urteilsziel~~ink~~onsistente Hinweise zu einer systematischen Verarbeitung führen und somit die Akkuratheit erhöhen.

Die Ergebnisse sowohl von Experiment 2 als auch von Experiment 3 deuten darauf hin, dass Lehrpersonen eine höhere Trefferquote erzielen bzw. eine höhere Trefferwahrscheinlichkeit aufweisen, nachdem sie einen urteilsziel~~ink~~onsistenten Hinweisreiz erhalten haben. Dies kann dadurch erklärt werden, dass Lehrpersonen – wie in Experiment 1 gezeigt – die Schüleraussagen systematischer verarbeiten und – wie Experiment 3 zeigen konnte – verbale Informationen für die Bildung ihres Glaubwürdigkeitsurteils nutzen. Diese Hypothese wird nur für Lehrkräfte mit einem Urteilsziel „wahr“ unterstützt, was durch den bereits weiter oben diskutierten Haupteffekt des Hinweisreizes auf das Glaubwürdigkeitsurteil erklärt werden kann. Inwiefern Lehrkräfte in Abhängigkeit des Verarbeitungsmodus verbale oder nonverbale Hinweise nutzen, wurde in dieser Arbeit nicht systematisch untersucht. Die Annahme, dass eine systematische Informationsverarbeitung zu einer Nutzung von verbalen Informationen bzw. Merkmalen des Nachrichteninhalts führt, wurde aus Befunden von Reinhard und Sporer (2008, 2010) abgeleitet.

8.1.3 Auswertung der Daten mittels Mehrebenenregressionen

Mehrebenenregressionen haben den großen Vorteil, dass nicht nur der Einfluss von Variablen auf Individualebene getestet werden kann, sondern gleichzeitig auch Variablen der Gruppenebene berücksichtigt werden können. Weiterhin können auch sog. Cross-Level-Interaktionen gebildet werden, mit deren Hilfe Hypothesen getestet werden können, bei denen der Einfluss von Variablen auf Individualebene in Abhängigkeit von Gruppenvariablen überprüft

werden kann. In der vorliegenden Arbeit sind die Faktoren „Urteilsvermeidung“ und „Hinweisreiz“ auf Gruppenebene angesiedelt; die „Gruppen“ sind hierbei die Lehrpersonen. Auf Individualebene befinden sich die zu beurteilenden Schüler mit „Länge der Aussagen“. Die Urteile über die Schüler sind in Lehrpersonen geschachtelt, wobei jede Lehrkraft acht Schüleraussagen beurteilte. Da es hier zu Überschneidungen kam und die in Sets organisierten Schüleraussagen sich überschneiden, handelt es sich um ein nicht-genestetes Modell.

In Experiment 3 wurde getestet, inwiefern die Urteilsvermeidung und der Hinweisreiz in Abhängigkeit der Antwortlänge die Wahrscheinlichkeit, die Aussage als „wahr“ einzustufen, beeinflussen. Anhand einer ANOVA hätte diese Hypothese mit den vorliegenden Daten überprüft werden können, indem eine nachträgliche Kategorisierung der Schüleraussagen in z.B. eher lang und eher kurz anhand eines Mediansplits vorgenommen worden wäre. Allerdings führt eine derartige Aggregation der Daten wie bereits oben erwähnt zu ungenaueren Schätzungen. Eine andere Möglichkeit, den Einfluss der Schülervariable in einer ANOVA zu berücksichtigen, wäre eine systematische Variation der Antwortlänge bei Erstellung des Videomaterials.

Ein weiterer Vorteil von Mehrebenenregression ist, dass Einflüsse von metrischen Variablen (z.B. des Alters der Probanden) bei Messwiederholungen getestet werden können, und dass die Variablen nicht aggregiert werden müssen (z.B. Einzelurteile nicht zu einem Summenscore zusammengefasst werden müssen). Die Daten von Experiment 3 wurden zum einen varianzanalytisch und zum anderen mehrebenenregressionsanalytisch

ausgewertet. Die Befunde beider Analysemethoden zeigen ähnliche Ergebnisse bezüglich der vorhergesagten Effekte. Zusätzlich werden durch die Varianzanalyse Effekte erzeugt, die regressionsanalytisch nicht aufgedeckt wurden. Dies deutet darauf hin, dass die Schätzung der Parameter in einem Messwiederholungsdesign akkurater sind, wenn die Auswertung der Daten regressionsanalytisch erfolgt.

Ein weiterer Vorteil, der für die vorliegende Arbeit von untergeordneter Bedeutung ist, ist die Tatsache, dass bei Mehrebenenanalysen hierarchische Strukturen von Daten berücksichtigt werden und dadurch Fehlinterpretationen vermieden werden können (vgl. Hox, 2010, und Twisk, 2011).

Aufgrund der genannten Vorteile scheint eine Auswertung von Forschungsdaten, bei denen es nachweislich einen Effekt des zu beurteilenden Objektes oder der zu beurteilenden Person auf das (Glaubwürdigkeits-)Urteil gibt, sinnvoll. Auch bei Designs, bei denen der Urteiler mehrere Urteile abgibt (im Sinne einer Messwiederholung), scheint eine Mehrebenenanalyse sinnvoll zu sein, da hier abhängige Variablen nicht aggregiert werden müssen.

8.2 Praktische Implikationen

Praktische Implikationen können zum einen aus der Art der Manipulation des Urteilsziels gezogen werden, aber auch aus den Befunden zum Verarbeitungsmodus und zur Schülervariable „Antwortlänge“ abgeleitet werden.

Die Manipulation des Urteilsziels basiert auf dem grundlegenden Bedürfnis, Aufwand zu minimieren und Ertrag zu maximieren. Eine Urteilsverzerrung wurde dadurch erzielt, dass die Lehrpersonen in

Abhängigkeit ihres Urteils zusätzlichen Aufwand im Sinne einer zusätzlichen, ausführlichen Begründung betreiben müssen. Um diesen zusätzlichen Aufwand zu umgehen, werden Urteile gefällt, die keine aufwendigen Konsequenzen nach sich ziehen. Dieses Szenario lässt sich leicht auf den Schulalltag übertragen: Hat ein Lehrer den Verdacht, dass einer seiner Schüler getäuscht hat (z.B. bei einer Klassenarbeit einen Spickzettel benutzt hat), so hat der Lehrer zum einen die Option, dem Verdacht nachzugehen und den Schüler zu befragen, und zum anderen, die Sache auf sich beruhen zu lassen und sich vorzunehmen, bei der nächsten Klassenarbeit den Schüler genauer zu beobachten. Die Befragung eines Schülers an sich ist bereits mit mehr Aufwand verbunden als ihn nicht zu befragen. Eine praktische Implikation könnte sein, jeder Lehrkraft nicht nur Zeit für Vermittlung des Lehrstoffs und Vor- und Nachbereitung des Unterrichts einzuräumen, sondern auch Zeit, um regelmäßig mit Schülern Gespräche führen zu können o.Ä. Entscheidet sich der Lehrer für eine Befragung, so stehen wiederum zwei Möglichkeiten zur Wahl. Entweder glaubt er dem Schüler, was vermutlich keine Konsequenzen nach sich zieht. Oder er schenkt den Aussagen des Schülers keinen Glauben, was sicherlich mit aufwändigen Konsequenzen verbunden ist, z.B. die Weiterleitung an den Direktor, die Eltern oder den Vertrauenslehrer. Eine praktische Implikation, die daraus abgeleitet werden könnte, ist es, der Lehrkraft mehr konsequenzenfreien Handlungsspielraum einzuräumen und z.B. bestimmtes Fehlverhalten von Schülern nicht an Dritte weiterleiten zu müssen.

Anhand der Befunde bezüglich des Effekts des Verarbeitungsmodus auf die Akkuratheit von Glaubwürdigkeitsurteilen in Experiment 2 und 3 lassen sich

praktische Implikationen ableiten, die auf eine Erhöhung der Motivation, Informationen systematisch zu verarbeiten, abzielen. Eine systematische Verarbeitung ist wahrscheinlicher, wenn Personen nicht unter kognitiver Belastung (cognitive load, eigene Übersetzung, Paas, Renkl, & Sweller, 2003) stehen (vgl. Reinhard & Sporer, 2008). Kognitive Belastung entsteht z.B. unter Stress oder bei der gleichzeitigen Bearbeitung mehrerer Aufgaben (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). Somit sollten Urteile über die Glaubwürdigkeit von Schülern unter möglichst geringer kognitiver Belastung gefällt werden (z.B. nicht während der Unterrichtszeit oder am Ende des Schultages).

Neben den Implikationen in Bezug auf den Urteiler, kann aus Experiment 3 eine Implikation in Bezug auf den Beurteilten abgeleitet werden. Aufgrund des Einflusses der Antwortlänge auf das Urteil, wäre es sinnvoll, bei der Befragung mehrerer Schüler darauf zu achten, dass die Antwortzeit möglichst konstant gehalten wird. Dies kann z.B. dadurch erreicht werden, dass den Schülern mitgeteilt wird, dass sie eine gewisse Zeit zum Antworten auf die Frage haben, diese auch nutzen sollten, aber dass keine Zeit für weitere Erklärungen sei.

8.3 Ausblick

Wie bereits zu Anfang der Abschlussdiskussion angemerkt, wurde in der vorliegenden Arbeit nicht untersucht, inwiefern es sich bei Urteilszielen tatsächlich um Ziele handelt. Laut Uleman (1996, 1999) können Ziele als nicht direkt zugänglich und unbewusst angesehen werden. Eine Möglichkeit zur Erfassung von unbewussten Zielen ist der Implicit Association Test (IAT;

Greenwald et al., 1998), mit dem klassischerweise implizite Einstellungen gemessen werden. Beim IAT müssen Probanden keine Fragen beantworten, sondern werden gebeten, Wörter oder Bilder auf einem Computer vorgegebenen Kategorien zuzuordnen. Gemessen wird dann das Ausmaß, indem zwei Konzepte miteinander assoziiert werden. Implizite Einstellungen werden erfasst, indem verglichen wird, wie schnell oder langsam bestimmte Hinweise mit negativen oder positiven Konzepten assoziiert werden. So könnte z.B. verglichen werden, wie schnell Lehrkräfte mit einem Urteilsziel „wahr“ Schülerhinweise (wie das Bild eines Schülers) mit Konzepten wie Ehrlichkeit oder Betrug assoziieren.

Die genaue Erforschung, inwiefern Personen, die ein Urteilsziel verfolgen, in Abhängigkeit vom Verarbeitungsmodus verbale oder nonverbale Hinweise nutzen, wäre ebenfalls interessant. So konnte Experiment 3 zwar zeigen, dass Probanden der Bedingung „Vermeidung Urteil erfunden“ durch die Länge der Antwort beeinflusst werden, wenn sie einen inkonsistenten Hinweisreiz erhalten haben; eine systematische Untersuchung der Nutzung verschiedener Informationen erfolgte jedoch nicht. Eine Möglichkeit der systematischen Untersuchung könnte in Anlehnung an Reinhard und Sporer (2008, 2010) durch eine systematische Variation verschiedener Indikatoren (z.B. Antwortlänge) in den vorgespielten Videosequenzen sein. So kann angenommen werden, dass Personen mit einem urteilsziel~~ink~~onsistenten (vs. urteilszielkonsistenten) Hinweisreiz den Aussagen anderer mehr Glaubwürdigkeit zuschreiben, wenn deren Aussagen überdurchschnittlich lang

sind. Durchschnittlich lange Antworten sollten das Glaubwürdigkeitsurteil dagegen nicht beeinflussen.

Weiterhin könnte in zukünftigen Experimenten die Stichprobengröße erhöht werden, um mehr als eine einzelne Schülervariable in den Analysen berücksichtigen zu können. Dies könnte dadurch erreicht werden, dass der Ausschluss von Probanden aus den Analysen reduziert wird und die Erhebung vor Ort in einem Labor stattfindet. So sind die Erhebungsbedingungen von Online-Befragungen im Vergleich zu Laborstudien nicht kontrollierbar und mögliche Einflussfaktoren wie z.B. Lärm oder Internetverbindung können nicht konstant gehalten oder in den Analysen als Kontrollvariablen mit aufgenommen werden (vgl. Sue & Ritter, 2011).

9 Literatur

- Aronson, E., Wilson, T. D., & Akert, R. (2008). *Sozialpsychologie*. München: Pearson Studium.
- Asch, S. E. (1956). Studies of independence and conformity. *Psychological Monographs*, 70(9), 1-70. doi: 10.1037/h0093718
- Bohner, G., Rank, S., Reinhard, M. A., Einwiller, S., & Erb, H. P. (1998). Motivational determinants of systematic processing: expectancy moderates effects of desired confidence on processing effort. *European Journal of Social Psychology*, 28(2), 185-206. doi: 10.1002/(Sici)1099-0992(199803/04)28:2<185::Aid-Ejsp864>3.0.Co;2-O
- Bond, C. F. (2006). A world of lies. *Journal of Cross-Cultural Psychology*, 37, 60-74. doi: 10.1177/0022022105282295
- Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, 10(3), 214-234. doi: 10.1207/s15327957pspr1003_2
- Bond, G. D., Malloy, D. M., Thompson, L. A., Arias, E. A., & Nunn, S. N. (2004). Post-probe decision making in a prison context. *Communication Monographs*, 71(3), 269-285. doi: 10.1080/0363452042000288328
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation*. Heidelberg: Springer. doi: 10.1007/978-3-540-33306-7
- Brehm, S. S., Kassin, S. M., & Fein, S. (2002). *Social psychology*. Boston: Houghton Mifflin Company.
- Cacioppo, J. T., Kao, C. F., Petty, R. E., & Rodriguez, R. (1986). Central and peripheral routes to persuasion - An individual difference perspective.

- Journal of Personality and Social Psychology*, 51(5), 1032-1043. doi: 10.1037/0022-3514.51.5.1032
- Cacioppo, J. T., & Petty, R. E. (1982). The Need for Cognition. *Journal of Personality and Social Psychology*, 42(1), 116-131. doi: 10.1037/0022-3514.42.1.116
- Cacioppo, J. T., & Petty, R. E. (1984). The Elaboration Likelihood Model of Persuasion. *Advances in Consumer Research*, 11(1), 673-675.
- Cacioppo, J. T., Petty, R. E., & Morris, K. J. (1983). Effects of Need for Cognition on message evaluation, recall, and persuasion. *Journal of Personality and Social Psychology*, 45(4), 805-818. doi: 10.1037/0022-3514.45.4.805
- Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39(5), 752-766. doi: 10.1037//0022-3514.39.5.752
- Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion process. In J. S. Uleman (Hrsg.), *Unintended thought* (S. 212-252). New York: Guilford Press.
- Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing - Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology*, 66(3), 460-473. doi: 10.1037/0022-3514.66.3.460

- Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. New York: Guilford Press.
- Chen, S. X., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Hrsg.), *Dual-process theories in social psychology* (S. 73-96). New York: Guilford Press.
- Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, 44(1), 20-33. doi: 10.1037/0022-3514.44.1.20
- Davis, S. F., Grover, C. A., Becker, A. H., & McGregor, L. N. (1992). Academic dishonesty - Prevalence, determinants, techniques, and punishments. *Teaching of Psychology*, 19(1), 16-20. doi: 10.1207/s15328023top1901_3
- Davison, A. C. (2003). *Statistical models*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511815850
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129(1), 74-118. doi: 10.1037/0033-2909.129.1.74
- Dickhäuser, O., Reinhard, M. A., & Marksteiner, T. (2012). Accurately detecting students' lies regarding relational aggression by correctional instructions. *Educational Psychology*, 32(2), 257-271. doi: 10.1080/01443410.2011.645271
- Ekman, P., O'Sullivan, M., & Frank, M. G. (1999). A few can catch a liar. *Psychological Science*, 10(3), 263-266. doi: 10.1111/1467-9280.00147

- Elliot, A. J., & Fryer, J. W. (2008). The goal construct in psychology. In J. Y. Shah & W. L. Gardner (Hrsg.), *Handbook of motivation science* (S. 235-250). New York: Guilford Press. doi: 10.1177/0011000007309979
- Evans, E. D., & Craig, D. (1990). Teacher and student perceptions of academic cheating in middle and senior High-Schools. *Journal of Educational Research*, 84(1), 44-52.
- Exline, R. V., Thibaut, J., Hickey, C. B., & Gumpert, P. (1970). Visual interaction in relation to Machiavellianism and an unethical act. In R. Christie, F. L. Geis & D. Berger (Hrsg.), *Studies in machiavellianism*. New York: Academic Press.
- Fabrigar, L. R., Priester, J. R., Petty, R. E., & Wegener, D. T. (1998). The impact of attitude accessibility on elaboration of persuasive messages. *Personality and Social Psychology Bulletin*, 24(4), 339-352. doi: 10.1177/0146167298244001
- Fahrmeir, L., Pigeot, I., Künstler, R., & Tutz, G. (2009). *Statistik: Der Weg zur Datenanalyse*: Springer DE.
- Field, A. (2009). *Discovering statistics using SPSS*. London: SAGE Publications.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London: SAGE Publications.
- Franklyn-Stokes, A., & Newstead, S. E. (1995). Undergraduate cheating: Who does what and why? *Studies in Higher Education*, 20(2), 159-172. doi: 10.1080/03075079512331381673

- Garrido, E., Masip, J., & Herrero, C. (2004). Police officers' credibility judgments: Accuracy and estimated ability. *International Journal of Psychology*, 39(4), 254-275. doi: 10.1080/00207590344000411
- Garst, J., & Bodenhausen, G. V. (1996). "Family values" and political persuasion: Impact of kin-related rhetoric on reactions to political campaigns. *Journal of Applied Social Psychology*, 26(13), 1119-1137. doi: 10.1111/j.1559-1816.1996.tb02288.x
- Gelman, A., & Hill, J. (2009). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.
- Giner-Sorolila, R., & Chaiken, S. (1997). Selective use of heuristic and systematic processing under defense motivation. *Personality and Social Psychology Bulletin*, 23(1), 84-97. doi: 10.1177/0146167297231009
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464-1480. doi: 10.1037/0022-3514.74.6.1464
- Hafer, C. L., Reynolds, K. L., & Obertynski, M. A. (1996). Message comprehensibility and persuasion: Effects of complex language in counterattitudinal appeals to laypeople. *Social Cognition*, 14(4), 317-337. doi: 10.1521/soco.1996.14.4.317
- Hartwig, M., & Bond Jr, C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin*, 137(4), 643-659. doi: 10.1037/a0023589

- Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression-formation. *Journal of Experimental Social Psychology*, 13(2), 141-154. doi: 10.1016/S0022-1031(77)80007-3
- Hox, J. J. (2010). *Multilevel analysis* (J. J. Hox Ed.). NY, New York: Routledge.
- Hubbell, A. P., Mitchell, M. M., & Gee, J. C. (2001). The relative effects of timing of suspicion and outcome involvement on biased message processing. *Communication Monographs*, 68(2), 115-132. doi: 10.1080/03637750128056
- Hull, C. (1943). *Principles of behavior*. New York: D. Appelton and Company.
- Hurst-Wagner, M., & Oswald, M. E. (2011). Mechanisms underlying response bias in deception detection. *Psychology, Crime & Law*, 18(8), 759-778. doi: 10.1080/1068316X.2010.550615
- Jäger, A. O. (2006). *Berliner Intelligenzstruktur-Test für Jugendliche Begabungs-und Hochbegabungsdiagnostik; BIS-HB*. Göttingen: Hogrefe.
- Köhnken, G. (1990). *Glaubwürdigkeit*. München: Psychologie Verlags Union.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480-498. doi: 10.1037/0033-2909.108.3.480
- Kunda, Z. (1999). *Social cognition: Making sense of people*. London: The MIT Press.
- Leippe, M. R., & Elkin, R. A. (1987). When motives clash - Issue involvement and response involvement as determinants of persuasion. *Journal of Personality and Social Psychology*, 52(2), 269-278. doi: 10.1037/0022-3514.52.2.269

- Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the "veracity effect". *Communication Monographs*, 66(2), 125-144. doi: 10.1080/03637759909376468
- Levine, T. R., Serota, K. B., Shulman, H., Clare, D. D., Park, H. S., Shaw, A. S., . . . Lee, J. H. (2011). Sender demeanor: Individual differences in sender believability have a powerful impact on deception detection judgments. *Human Communication Research*, 37(3), 377-403. doi: 10.1111/j.1468-2958.2011.01407.x
- Locke, E. A., Saari, L. M., Shaw, K. N., & Latham, G. P. (1981). Goal setting and task-performance - 1969-1980. *Psychological Bulletin*, 90(1), 125-152. doi: 10.1037/0033-2909.90.1.125
- Maheswaran, D., & Chaiken, S. (1991). Promoting systematic processing in low-motivation settings - Effect of incongruent information on processing and judgment. *Journal of Personality and Social Psychology*, 61(1), 13-25. doi: 10.1037/0022-3514.61.1.13
- Maio, G. R., Bell, D. W., & Esses, V. M. (1996). Ambivalence and persuasion: The processing of messages about immigrant groups. *Journal of Experimental Social Psychology*, 32(6), 513-536. doi: 10.1006/jesp.1996.0023
- Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: Police officers' ability to detect suspects' lies. *Journal of Applied Psychology*, 89(1), 137-149. doi: 10.1037/0021-9010.89.1.137
- Marksteiner, T., Dickhäuser, O., & Reinhard, M. A. (2012). Der Zusammenhang von Need for Cognition und Stereotypen bei der Beurteilung der

Glaubwürdigkeit von Schülern. *Psychologie in Erziehung Und Unterricht*, 59(1), 47-59.

Marksteiner, T., Plieninger, H., Reinhard, M. A., & Dickhäuser, O. (2013).

Biases in credibility judgments: Using logistic multi-level regressions for accounting for sender influences. Manuscript in Vorbereitung.

Marksteiner, T., & Reinhard, M.-A. (2013). *Cheating.* Manuskript, eingereicht zur Publikation.

Marksteiner, T., Reinhard, M. A., Dickhäuser, O., & Sporer, S. L. (2012). How do teachers perceive cheating students? Beliefs about cues to deception and detection accuracy in the educational field. *European Journal of Psychology of Education*, 27(3), 329-350. doi: 10.1007/s10212-011-0074-5

Marksteiner, T., Reinhard, M. A., Lettau, F., & Dickhäuser, O. (2013). Bullying, cheating, deceiving: Teachers' perception of deceitful situations at school. *Manuskript, eingereicht zur Publikation.*

Masip, J., Alonso, H., Garrido, E., & Anton, C. (2005). Generalized communicative suspicion (GCS) among police officers: Accounting for the investigator bias effect. *Journal of Applied Social Psychology*, 35(5), 1046-1066. doi: 10.1111/j.1559-1816.2005.tb02159.x

Masip, J., Garrido, E., & Herrero, C. (2006). Observers' decision moment in deception detection experiments: Its impact on judgment, accuracy, and confidence. *International Journal of Psychology*, 41(4), 304-319. doi: 10.1080/00207590500343612

- Masip, J., Garrido, E., & Herrero, C. (2009). Heuristic versus systematic processing of information in detecting deception: Questioning the truth bias. *Psychological Reports, 105*(1), 11-36. doi: 10.2466/pr0.105.1.11-36
- McAllister, D. W., Mitchell, T. R., & Beach, L. R. (1979). The contingency model for the selection of decision strategies: An empirical test of the effects of significance, accountability, and reversibility. *Organizational Behavior and Human Performance, 24*(2), 228-244. doi: 10.1016/0030-5073(79)90027-8
- Meissner, C. A., & Kassin, S. M. (2002). "He's guilty!": Investigator bias in judgments of truth and deception. *Law and Human Behavior, 26*(5), 469-480. doi: 10.1023/A:1020278620751
- Moreno, R., & Mayer, R. E. (2000). Engaging students in active learning: The case for personalized multimedia messages. *Journal of Educational Psychology, 92*(4), 724-733. doi: 10.1037/0022-0663.92.4.724
- OECD. (2009). PISA data analysis manual: SPSS (2nd ed.). Paris, France: Author. Retrieved from doi:10.1787/9789264056275-en
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*(4), 867-872. doi: 10.1016/j.jesp.2009.03.009
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1-4. doi: 10.1207/S15326985EP3801_1

- Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, 38(1), 63-71. doi: 10.1207/S15326985EP3801_8
- Pegels, E. M. (1997). *Mogeln und Moral*. Berlin: LIT Verlag.
- Petty, R. E., Goldman, R., & Cacioppo, J. T. (1981). Personal involvement as a determinant of argument-based persuasion. *Journal of Personality and Social Psychology*, 41(5), 847-855. doi: 10.1037/0022-3514.41.5.847
- Phares, E. J., & Chaplin, W. F. (1997). *Introduction to personality*. New York: Longman.
- Pizarro, D. A. (2010). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. *Paper presented at the Herzliya Symposia on Personality and Social Psychology*.
- Reinhard, M. A. (2010). Need for Cognition and the process of lie detection. *Journal of Experimental Social Psychology*, 46(6), 961-971. doi: 10.1016/j.jesp.2010.06.002
- Reinhard, M. A., Dickhäuser, O., Marksteiner, T., & Sporer, S. L. (2011). The case of Pinocchio: Teachers' ability to detect deception. *Social Psychology of Education*, 14(3), 299-318. doi: 10.1007/s11218-010-9148-5
- Reinhard, M. A., & Sporer, S. L. (2008). Verbal and nonverbal behaviour as a basis for credibility attribution: The impact of task involvement and cognitive capacity. *Journal of Experimental Social Psychology*, 44(3), 477-488. doi: 10.1016/j.jesp.2007.07.012

- Reinhard, M. A., & Sporer, S. L. (2010). Content versus source cue information as a basis for credibility judgments: The impact of task involvement. *Social Psychology, 41*(2), 93-104. doi: 10.1027/1864-9335/a000014
- Reinhard, M. A., Sporer, S. L., Scharmach, M., & Marksteiner, T. (2011). Listening, not watching: Situational familiarity and the ability to detect deception. *Journal of Personality and Social Psychology, 101*(3), 467-484. doi: 10.1037/a0023726
- Smith, S. M., & Petty, R. E. (1996). Message framing and persuasion: A message processing analysis. *Personality and Social Psychology Bulletin, 22*(3), 257-268. doi: 10.1177/0146167296223004
- Sue, V. M., & Ritter, L. A. (2011). *Conducting online surveys*: Sage Publications, Incorporated.
- Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin, 121*(3), 371-394. doi: 10.1037/0033-2909.121.3.371
- Taylor, S. E. (1981). The interface of cognitive and social psychology. *Cognition, social behavior, and the environment, 189-211*.
- Terhart, E. (2005). Standards für die Lehrerbildung - Ein Kommentar. *Zeitschrift für Pädagogik, 51*(2), 275-279.
- Turco, R. M. (1996). Self-referencing, quality of argument, and persuasion. *Current Psychology, 15*(3), 258-276. doi: 10.1007/BF02686883
- Twisk, J. W. R. (2011). *Applied multilevel analysis*. Cambridge: Cambridge University Press.

- Uleman, J. S. (1996). Integrative views of motivation, cognition, and emotion - Spaulding, WD. *Contemporary Psychology*, 41(2), 109-111.
- Uleman, J. S. (1999). Spontaneous versus intentional inferences in impression formation. In S. Chaiken & Y. Trope (Hrsg.), *Dual-process theories in social psychology* (S. 141-160). New York: Guilford Press.
- Vrij, A. (2008). *Detecting lies and deceit*. Chichester, West Sussex: John Wiley and Sons.
- Vrij, A., Akehurst, L., Brown, L., & Mann, S. (2006). Detecting lies in young children, adolescents and adults. *Applied Cognitive Psychology*, 20(9), 1225-1237. doi: 10.1002/acp.1278
- Vrij, A., Granhag, P. A., & Porter, S. (2010). Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychological Science in the Public Interest*, 11(3), 89-121. doi: 10.1177/1529100610390861
- West, T. V., & Kenny, D. A. (2011). The Truth and Bias Model of Judgment. *Psychological Review*, 118(2), 357-378. doi: 10.1037/a0022936
- Winer, B. J. (1971). Statistical reasoning in psychology and education. *Contemporary Psychology*, 16(2), 84-85.
- Wood, W. (2000). Attitude change: Persuasion and social influence. *Annual Review of Psychology*, 51, 539-570. doi: 10.1146/annurev.psych.51.1.539

Inhaltsverzeichnis

I.	Anhang zu Pilotstudie 1 und 2 und Experiment 1	162
a.	Ausschnitt Brief	162
b.	Gespräch	162
c.	Abbildung Instruktion.....	164
II.	Anhang zu Experiment 2	164
a.	Syntax Simple-Effect-Analysen.....	164
III.	Anhang zu Experiment 3	164
a.	Erläuterung und Art der Erfassung der objektiven Einschätzungen	164
b.	Syntax Datenvorbereitung für Analysen in R	167
c.	Syntax Simple-Effect-Analysen.....	177

I. Anhang zu Pilotstudie 1 und 2 und Experiment 1

a. Ausschnitt Brief

„Lieber Markus,
Mh na du arme alte Ratte? Um dich in deiner
Unsicherheit zu bestätigen will ich dir nun einen
Brief schreiben, um dich richtig zum Heuten
zu bringen. Alles fresh? Wahrscheinlich nicht, wenn
ich an deine alltägliche Hygiene denke... Hahah...
Du bist doch wirklich zu dumm. Obwohl du
dir jeden Tag was ...

b. Gespräch

Lehrer:

„Ich habe hier einen Brief vor mir liegen, der äußerst negative Äußerungen über Markus enthält. Er hat mir erzählt, dass er diesen Brief in seiner Tasche gefunden hat, die er in der vierten Stunde unbeaufsichtigt im Aufenthaltsraum hat liegen lassen. Markus ist ziemlich verletzt über diesen Brief. Warst du am Verfassen dieses Briefes beteiligt?“

Schüler S.:

„Nein. Ich weiß gar nichts von nem Brief ...“

Lehrer:

„Kannst du mir sagen, wo genau du dich in der vierten Stunde aufgehalten hast?“

Schüler S.:

„Ich war die ganze Zeit aufm Boden gesessen, so 6, 7 Meter vom Aufenthaltsraum weg. Wir hatten ja ne Freistunde in der vierten.“

Lehrer:

„Und was hast du gemacht?“

Schüler S.:

„Musik gehört und in dem Buch gelesen, dass wir für Deutsch lesen sollen.“

Lehrer:

„Und wo waren die anderen?“

Schüler S.:

„Ein paar waren neben mir aufm Boden gesessen. Die anderen weiß ich nich ...“

Lehrer:

„Was genau hast du und die anderen denn gemacht?“

Schüler S.:

„Wir haben uns unterhalten, über Kinofilme, Hobbies und die Schule. Also, hauptsächlich über Kinofilme, weil das am interessantesten war.“

Lehrer:

„Welche Erklärung hast du dafür, dass Markus einen solchen Brief in seiner Tasche gefunden hat?“

Schüler S.:

„Kann ich mir nicht erklären, weiß ich nicht, wie der da hingekommen ist, weil wir uns wirklich nur über Kinofilme, ja, hauptsächlich eigentlich über Kinofilme unterhalten haben ...“

Lehrer:

„Ihr wart ja eine ganze Weile vor dem Aufenthaltsraum gesessen. In der Zeit hättet ihr ja einen Brief verfassen können und in Markus Tasche stecken können, oder?“

Schüler S.:

„Klar hätten wir Zeit gehabt. Aber es gäb ja keinen Grund das zu tun. Also, zumindest nicht aus meiner Sicht.“

Lehrer:

„In dem Brief steht: 'Alles fresh? Wahrscheinlich nicht, wenn ich an deine alltägliche Hygiene denke. Du bist doch wirklich zu dumm.' Was muss in jemandem vorgehen, der so etwas schreibt?“

Schüler S.:

„Ja, trifft ja aber nicht zu. Also, er sieht ja ganz normal aus. Deswegen hätt ich ja keinen Kopp gehabt, so einen Brief zu verfassen.“

Lehrer:

„Also ich denke, dass Markus doch sehr verletzt wurde. Ich denke, es ist an der Zeit, dass derjenige, der den Brief verfasst hat, Verantwortung übernimmt und zugibt, was er getan hat!“

Schüler S.:

„Hm. Seh ich auch so. Aber ich geb ja nur was zu, was ich auch gemacht hab ... Und das hab ich nicht getan ... Ganz definitiv nicht und wenn ich mir da sicher sein kann, dann geb ich das ja nicht zu.“

c. Abbildung Instruktion

Instruktion für „Urteilsziel wahr“:

Die Aussagen des Schülers halte ich für ...?

erlogen	1	2	3	4	5		6	7	8	9	10	wahr
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
	Fallbericht verfassen						keinen Fallbericht verfassen					

Instruktion für „Urteilsziel erfunden“:

Die Aussagen des Schülers halte ich für ...?

erlogen	1	2	3	4	5		6	7	8	9	10	wahr
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
	keinen Fallbericht verfassen						Fallbericht verfassen					

II. Anhang zu Experiment 2

a. Syntax Simple-Effect-Analysen

*simple effects analysis.

```
GLM corr_allpro by Urteilsziel Hinweisreiz
/emmeans = tables(Urteilsziel*Hinweisreiz)compare(Urteilsziel).
```

```
GLM corr_allpro by Urteilsziel Hinweisreiz
/emmeans = tables(Urteilsziel*Hinweisreiz)compare(Hinweisreiz).
```

III. Anhang zu Experiment 3

a. Erläuterung und Art der Erfassung der objektiven

Einschätzungen

Antwortlänge (response length, 1)

Länge oder Dauer der Botschaft des Senders. Dauer in Sekunden:__

Details (details, 4)

Grad, zu welchem die Botschaft Details enthält, wie z.B. Beschreibungen von Personen, Handlungen, Orten, Objekten, Ereignissen und zeitliche Aspekte. Anzahl: _

Plausibilität (plausibility, 12)

unplausibel						plausibel
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1	2	3	4	5	6	7

Logische Struktur (logical structure: “consistency” and “coherence”, 13)

Konsistenz und Kohärenz der Botschaft; "Sammlung von verschiedenen und unabhängigen Details, welche insgesamt eine kohärente Darstellung von aufeinanderfolgenden Ereignissen sind."

unlogisch						logisch
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1	2	3	4	5	6	7

Blickabwendung (wo sitzt Interviewer?) (gaze aversion, 28)

Sprecher sieht weg oder wendet den Blick ab

→ Vorkommen pro 10 Sekunden: ja/nein

10s	
20s	
30s	
40s	
50s	
60s	
70s	
80s	
90s	
100s	
110s	
120s	
130s	
140s	
150s	
160s	
170s	
180s	
190s	
200s	
210s	
220s	

230s	
240s	

Nervosität (nervous, tense (overall), 61)

Sprecher erscheint nervös und angespannt. Sprecher zeigt Körperbewegungen, die ihn nervös erscheinen lassen.

nicht nervös						nervös
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1	2	3	4	5	6	7

b. Syntax Datenvorbereitung für Analysen in R

****** Neue Variablen v01 bis v08 bilden und Wert auf null setzen.

```
compute v01 = 0.  
compute v02 = 0.  
compute v03 = 0.  
compute v04 = 0.  
compute v05 = 0.  
compute v06 = 0.  
compute v07 = 0.  
compute v08 = 0.  
EXECUTE.
```

****** Für jedes Set festlegen, welcher Schüler beurteilt wurde.

***** Set 1.

```
if (set = 1) v01 = 28.  
if (set = 1) v02 = 03.  
if (set = 1) v03 = 11.  
if (set = 1) v04 = 15.  
if (set = 1) v05 = 44.  
if (set = 1) v06 = 22.  
if (set = 1) v07 = 53.  
if (set = 1) v08 = 25.  
EXECUTE.
```

***** Set 2.

```
if (set = 2) v01 = 54.  
if (set = 2) v02 = 62.  
if (set = 2) v03 = 29.  
if (set = 2) v04 = 37.  
if (set = 2) v05 = 58.  
if (set = 2) v06 = 60.  
if (set = 2) v07 = 55.  
if (set = 2) v08 = 83.  
EXECUTE.
```

***** Set 3.

```
if (set = 3) v01 = 68.  
if (set = 3) v02 = 84.  
if (set = 3) v03 = 21.  
if (set = 3) v04 = 43.  
if (set = 3) v05 = 76.  
if (set = 3) v06 = 90.  
if (set = 3) v07 = 85.  
if (set = 3) v08 = 95.  
EXECUTE.
```


* Set 4.

```
if (set = 4) v01 =44.
if (set = 4) v02 =22.
if (set = 4) v03 =53.
if (set = 4) v04 =25.
if (set = 4) v05 =54.
if (set = 4) v06 =62.
if (set = 4) v07 =29.
if (set = 4) v08 =37.
EXECUTE.
```

* Set 5.

```
if (set = 5) v01 =58.
if (set = 5) v02 =60.
if (set = 5) v03 =55.
if (set = 5) v04 =83.
if (set = 5) v05 =68.
if (set = 5) v06 =84.
if (set = 5) v07 =21.
if (set = 5) v08 =43.
EXECUTE.
```

* Set 6.

```
if (set = 6) v01 =76.
if (set = 6) v02 =90.
if (set = 6) v03 =85.
if (set = 6) v04 =95.
if (set = 6) v05 =28.
if (set = 6) v06 =03.
if (set = 6) v07 =11.
if (set = 6) v08 =15.
EXECUTE.
```

** Neue Variable status01 bis status08 bilden und definieren, welchen Status die Schülersage hat.

*Status Video: 1 = Wahrheit, 2 = Lüge.

```
compute status01 = 1.
compute status02 = 1.
compute status03 = 2.
compute status04 = 2.
compute status05 = 1.
compute status06 = 1.
compute status07 = 2.
compute status08 = 2.
EXECUTE.
```

```

** Neue Variable geschlecht01 bis geschlecht08 bilden und definieren, welches Geschlecht der Schüler hat.
*Status Video: Schüler Geschlecht.
compute geschlecht01 = 1.
compute geschlecht02 = 1.
compute geschlecht03 = 1.
compute geschlecht04 = 1.
compute geschlecht05 = 2.
compute geschlecht06 = 2.
compute geschlecht07 = 2.
compute geschlecht08 = 2.
EXECUTE.

if (set > 3) geschlecht01 = 2.
if (set > 3) geschlecht02 = 2.
if (set > 3) geschlecht03 = 2.
if (set > 3) geschlecht04 = 2.
if (set > 3) geschlecht05 = 1.
if (set > 3) geschlecht06 = 1.
if (set > 3) geschlecht07 = 1.
if (set > 3) geschlecht08 = 1.
EXECUTE.

** Neue Variable id bilden und Fallnummer zuweisen.
compute id = $casenum.
EXECUTE.

**metrisches Glaubwürdigkeitsurteil.
*Urteil01.
RECODE cred11 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) INTO cred01.
EXECUTE.

RECODE cred21 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) INTO cred01.
EXECUTE.

RECODE cred31 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) INTO cred01.
EXECUTE.

RECODE cred41 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) INTO cred01.
EXECUTE.

RECODE cred51 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) INTO cred01.
EXECUTE.

RECODE cred61 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) INTO cred01.
EXECUTE.

```

*Urteil 02.

RECODE cred12 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred02.
EXECUTE.

RECODE cred22 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred02.
EXECUTE.

RECODE cred32 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred02.
EXECUTE.

RECODE cred42 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred02.
EXECUTE.

RECODE cred52 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred02.
EXECUTE.

RECODE cred62 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred02.
EXECUTE.

*Urteil03.

RECODE cred13 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred03.
EXECUTE.

RECODE cred23 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred03.
EXECUTE.

RECODE cred33 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred03.
EXECUTE.

RECODE cred43 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred03.
EXECUTE.

RECODE cred53 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred03.
EXECUTE.

RECODE cred63 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred03.
EXECUTE.

*Urteil04.

RECODE cred14 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred04.
EXECUTE.

RECODE cred24 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred04.
EXECUTE.

RECODE cred34 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred04.
EXECUTE.

RECODE cred44 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred04.
EXECUTE.

RECODE cred54 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred04.
EXECUTE.

RECODE cred64 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred04.
EXECUTE.

*Urteil05.

RECODE cred15 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred05.
EXECUTE.

RECODE cred25 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred05.
EXECUTE.

RECODE cred35 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred05.
EXECUTE.

RECODE cred45 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred05.
EXECUTE.

RECODE cred55 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred05.
EXECUTE.

RECODE cred65 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred05.
EXECUTE.

*Urteil06.

RECODE cred16 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred06.
EXECUTE.

RECODE cred26 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred06.
EXECUTE.

RECODE cred36 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred06.
EXECUTE.

RECODE cred46 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred06.
EXECUTE.

RECODE cred56 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred06.
EXECUTE.

RECODE cred66 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred06.
EXECUTE.

*Urteil07.

RECODE cred17 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred07.
EXECUTE.

RECODE cred27 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred07.
EXECUTE.

RECODE cred37 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred07.
EXECUTE.

RECODE cred47 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred07.
EXECUTE.

RECODE cred57 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred07.
EXECUTE.

RECODE cred67 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred07.
EXECUTE.

*Urteil08.

RECODE cred18 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred08.
EXECUTE.

RECODE cred28 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred08.
EXECUTE.

RECODE cred38 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred08.
EXECUTE.

RECODE cred48 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred08.
EXECUTE.

RECODE cred58 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred08.
EXECUTE.

RECODE cred68 (1=1) (2=2) (3=3) (4=4) (5=5) (6=6) (7=7) (8=8) (9=9) (10=10) **INTO** cred08.
EXECUTE.

**binäres Glaubwürdigkeitsurteil.

*Urteil01bi.

RECODE truelie11 (1=1) (2=2) **INTO** truelie01.
EXECUTE.

RECODE truelie21 (1=1) (2=2) **INTO** truelie01.
EXECUTE.

RECODE truelie31 (1=1) (2=2) **INTO** truelie01.
EXECUTE.

RECODE truelie41 (1=1) (2=2) **INTO** truelie01.
EXECUTE.

RECODE truelie51 (1=1) (2=2) **INTO** truelie01.
EXECUTE.

RECODE truelie61 (1=1) (2=2) **INTO** truelie01.
EXECUTE.

*Urteil02bi.

RECODE truelie12 (1=1) (2=2) **INTO** truelie02.
EXECUTE.

RECODE truelie22 (1=1) (2=2) **INTO** truelie02.
EXECUTE.

RECODE truelie32 (1=1) (2=2) **INTO** truelie02.
EXECUTE.

RECODE truelie42 (1=1) (2=2) **INTO** truelie02.
EXECUTE.

RECODE truelie52 (1=1) (2=2) **INTO** truelie02.
EXECUTE.

RECODE truelie62 (1=1) (2=2) **INTO** truelie02.
EXECUTE.

*Urteil03bi.

RECODE truelie13 (1=1) (2=2) **INTO** truelie03.
EXECUTE.

RECODE truelie23 (1=1) (2=2) **INTO** truelie03.
EXECUTE.

RECODE truelie33 (1=1) (2=2) INTO truelie03.
EXECUTE.

RECODE truelie43 (1=1) (2=2) INTO truelie03.
EXECUTE.

RECODE truelie53 (1=1) (2=2) INTO truelie03.
EXECUTE.

RECODE truelie63 (1=1) (2=2) INTO truelie03.
EXECUTE.

*Urteil04bi.

RECODE truelie14 (1=1) (2=2) INTO truelie04.
EXECUTE.

RECODE truelie24 (1=1) (2=2) INTO truelie04.
EXECUTE.

RECODE truelie34 (1=1) (2=2) INTO truelie04.
EXECUTE.

RECODE truelie44 (1=1) (2=2) INTO truelie04.
EXECUTE.

RECODE truelie54 (1=1) (2=2) INTO truelie04.
EXECUTE.

RECODE truelie64 (1=1) (2=2) INTO truelie04.
EXECUTE.

*Urteil05bi.

RECODE truelie15 (1=1) (2=2) INTO truelie05.
EXECUTE.

RECODE truelie25 (1=1) (2=2) INTO truelie05.
EXECUTE.

RECODE truelie35 (1=1) (2=2) INTO truelie05.
EXECUTE.

RECODE truelie45 (1=1) (2=2) INTO truelie05.
EXECUTE.

RECODE truelie55 (1=1) (2=2) INTO truelie05.
EXECUTE.

RECODE truelie65 (1=1) (2=2) INTO truelie05.
EXECUTE.

*Urteil06bi.

RECODE truelie16 (1=1) (2=2) INTO truelie06.
EXECUTE.

RECODE truelie26 (1=1) (2=2) INTO truelie06.
EXECUTE.

RECODE truelie36 (1=1) (2=2) INTO truelie06.
EXECUTE.

RECODE truelie46 (1=1) (2=2) INTO truelie06.
EXECUTE.

RECODE truelie56 (1=1) (2=2) INTO truelie06.
EXECUTE.

RECODE truelie66 (1=1) (2=2) INTO truelie06.
EXECUTE.

*Urteil07bi.

RECODE truelie17 (1=1) (2=2) INTO truelie07.
EXECUTE.

RECODE truelie27 (1=1) (2=2) INTO truelie07.
EXECUTE.

RECODE truelie37 (1=1) (2=2) INTO truelie07.
EXECUTE.

RECODE truelie47 (1=1) (2=2) INTO truelie07.
EXECUTE.

RECODE truelie57 (1=1) (2=2) INTO truelie07.
EXECUTE.

RECODE truelie67 (1=1) (2=2) INTO truelie07.
EXECUTE.

*Urteil08bi.

RECODE truelie18 (1=1) (2=2) **INTO** truelie08.
EXECUTE.

RECODE truelie28 (1=1) (2=2) **INTO** truelie08.
EXECUTE.

RECODE truelie38 (1=1) (2=2) **INTO** truelie08.
EXECUTE.

RECODE truelie48 (1=1) (2=2) **INTO** truelie08.
EXECUTE.

RECODE truelie58 (1=1) (2=2) **INTO** truelie08.
EXECUTE.

RECODE truelie68 (1=1) (2=2) **INTO** truelie08.
EXECUTE.

** Datensatz umstrukturieren und Variablen zu Fällen machen.

VARSTOCASES

/MAKE videonr **FROM** v01 v02 v03 v04 v05 v06 v07 v08

/MAKE videostatus **FROM** status01 status02 status03 status04 status05 status06 status07 status08

/MAKE urteilkorrekt **FROM** corr1 corr2 corr3 corr4 corr5 corr6 corr7 corr8

/MAKE credible **FROM** cred01 cred02 cred03 cred04 cred05 cred06 cred07 cred08

/MAKE liettrue **FROM** truelie01 truelie02 truelie03 truelie04 truelie05 truelie06 truelie07 truelie08

/MAKE videogeschlecht **FROM** geschlecht01 geschlecht02 geschlecht03 geschlecht04 geschlecht05
geschlecht06 geschlecht07 geschlecht08

/INDEX=Index1(8)

/KEEP=id Urteilsziel Hinweisreiz Set Geschlecht Falschbeschuld Fehlverhaufd consistent truthbias

/NULL=KEEP.

** Variablen umkodieren.

*nach Rekodierung: 0 = erfunden, 1 = wahr.

RECODE liettrue (1=1) (2=0).
EXECUTE.

*nach Rekodierung: 0 = wahr, 1 = erfunden.

RECODE Urteilsziel (1=0) (2=1).

EXECUTE.

RECODE Hinweisreiz (1=0) (2=1).

EXECUTE.

RECODE Geschlecht (1=0) (2=1).

EXECUTE.

RECODE consistent (1=0) (2=1).

EXECUTE.

**Datensatz von Experiment 3 mit Schülervariablen mergen.

MATCH FILES /TABLE=*

/FILE='DatenSet3'

/BY videonr.

EXECUTE.

**Neue Variable Urteilsziel und Hinweisreiz; nach Rekodierung: 0 = erfunden, 1 = wahr.

RECODE Urteilsziel Hinweisreiz (1=0) (0=1) **INTO** Ziel_trli Cue_trli.

EXECUTE.

** Datensatz speichern in .dat-Format.

SAVE TRANSLATE OUTFILE='J:\Diss\Daten\Experiment3_Marksteiner_R.dat'

/TYPE=TAB

/MAP

/REPLACE

/FIELDNAMES

/CELLS=VALUES.

c. Syntax Simple-Effect-Analysen

*simple effects analysis.

GLM corr_allpro **by** Urteilsziel Hinweisreiz

/emmeans = **tables**(Urteilsziel*Hinweisreiz)**compare**(Urteilsziel).

GLM corr_allpro **by** Urteilsziel Hinweisreiz

/emmeans = **tables**(Urteilsziel*Hinweisreiz)**compare**(Hinweisreiz).